QUARTERLY OF APPLIED MATHEMATICS

Vol. VI OCTOBER, 1948 No. 3

CONJUGATE ACTION OF INVOLUTE HELICAL GEARS WITH PARALLEL OR INCLINED AXES*

BY

H. PORITSKY AND D. W. DUDLEY

General Electric Company

Introduction. A problem of interest in mechanics of machinery is the transformation of a uniform rotation about one axis A into a uniform rotation about another axis A'. This problem and the design of the gear surfaces give rise to interesting geometry. It is shown in the following (Theorems A and B) that such a transformation both for the case of parallel axes A, A' and non-parallel or skew axes A, A' may be accomplished by means of two involute helical surfaces to be described presently. The case of non-parallel axes is covered by Theorems A, B; that by parallel axes by Theorem C. It is further shown that where this transformation is possible, the ratio of velocities for any two involute helical surfaces is independent of the relative orientation of the two surfaces. This result is generalized in Theorems A′, B′, C′ by the inclusion of uniform translations; the ratio of rotational velocities, while constant, is then dependent also on the translational velocities. An alternative proof of Theorems A, B, C follows from Theorem D in which it is shown that by rotating an involute helicoid about its axis through an arbitrary angle one obtains a family of parallel surfaces. As a limiting case of Theorem C, if the axis A' is made to recede to infinity one obtains the mating action of a rotating involute helicoid and a plane which is undergoing a uniform translation; this is given in Theorem E.

Not only are these results of interest from point of view of *transmitting uniform rotation* by means of gears, (or transforming a rotation of a gear into a translation of a rack), but they are also of direct industrial interest from point of view of *generating* the *gear surfaces* in question. By designing the cutting tool (hob or shaving tool) so that the cutting edges lie on an involute helical surface, and providing proper rotation of the work and the tool, each about its axis, and proper translation of the tool parallel to the work axis, one can prove from the converse of the above theorems that the tool will generate an involute helical surface.†

The results of this paper will also serve to answer certain practical questions which constantly come up in gear manufacturing work. For instance, in hobbing helical involute gears the angle Σ between the axis of the hob and the axis of the gear is set to

*Received Sept. 5, 1947.

†It should be pointed out that if the generating tool is not a grinding wheel and possesses a finite number of cutting edges, it produces "scallops", and it is necessary from the practical standpoint to design and operate the tool so as to limit the size of scallops and thereby obtain a surface corresponding closely to the theoretical surfaces considered herein.

agree with the *design* lead angles of the two members. During the life of the hob, its outside diameter and lead angle change as its cutting surface is ground away in sharpening. Likewise in shaving, the tool changes size and its actual lead angle will vary from the design lead angle. Many gear engineers have questioned whether or not it was necessary to adjust the shaft angle in hobbing and shaving to agree with the actual lead angle of the tool. It has been thought that a discrepancy in the shaft angle of the cutting tool would produce a profile error in the part being machined.

In spiral gear drives, errors in machining the casings and gear parts will result in the gears being run at a shaft angle somewhat different than the design angle. Gear inspectors often question how accurately the shaft alignment must be held to maintain quiet running gears and adequate gear tooth contact.

These questions and many others can be answered by applying the results of this paper. Some of these practical applications to gears are discussed in Sec. 5.

While a general appreciation of the main properties of involute spur and helical gears is wide spread in industry, and some treatments of their properties may be found in the literature [see, for instance, the article by Nikola Trbojevich, "Problem of the Theoretically Correct Involute Hob", *Machinery*, **25**, 429–433 (1919)], the authors have not run across any existing comprehensive exposition of the inherently simple basic geometry and kinematics of these gears. Moreover, from personal industrial contacts, they are aware that the equations for finding the effect of changes in distance and angle on tooth shape, locus of contact, and backlash are either unknown, or not readily available. The present paper contains an exposition of the simple basic helical involute gear geometry and aims to provide equations from which the practical design and shop questions involving these gears may be answered.

1. Statement of results. By a *helical surface* is meant a surface generated by a curve C during a screw motion about an axis. Such a surface consists of helices generated by the individual points of C during the screw motion. By an *involute* is meant an involute of a proper circle known as the *base circle*, whose radius, the *base radius*, will be denoted by R_b . Finally, by an *involute helicoid* is meant a helical surface generated by an involute lying in a plane perpendicular to the axis of the screw motion, that is a helical surface such that a section of it by a plane perpendicular to the axis of the helical surface is an involute of a proper base circle.

In cylindrical coordinates r, θ, z, with the z-axis along the axis A of the screw motion and θ measured as a right-handed rotation from $y = 0$ about the positive z-axis, the equation of a helical surface is given by

$$r = f(z - C\theta), \tag{1.1}$$

where C represents the axial displacement corresponding to a rotation of one radian about the axis, and is positive for a right-handed screw, negative for a left-handed one.

To determine the equations of an involute helical surface, we start from the equations of an involute

$$\begin{aligned} x &= R_b [\cos \varphi_1 - (\varphi_2 - \varphi_1) \sin \varphi_1], \\ y &= R_b [\sin \varphi_1 + (\varphi_2 - \varphi_1) \cos \varphi_1], \end{aligned} \tag{1.2}$$

where φ_2 is the value of θ at the cusp of the involute on the *base circle* $r = R_b$, and φ_1 the value of θ at the point of contact of the tangent line with the base circle (Fig. 1)

By keeping φ_2 constant and varying φ_1 one obtains various involutes of the same base circle, the cusp corresponding to $\varphi_1 = \varphi_2$. For constant φ_1 and variable φ_2 the tangents to the base circle result. For $\varphi_1 > \varphi_2$, φ_1, φ_2 yield an orthogonal family of curves outside the base circle; similarly for $\varphi_1 < \varphi_2$. All involutes are congruent, any one being obtainable from any other one through a rotation about the origin of amount $\Delta\varphi_2$; in the

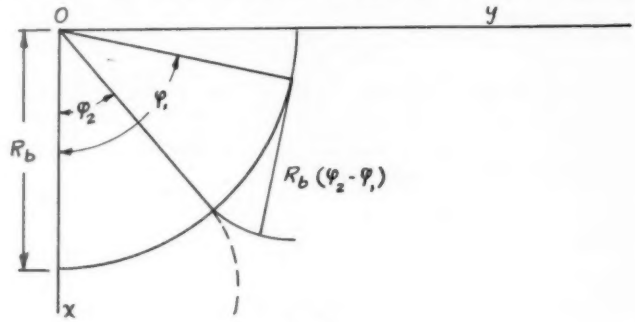

FIG. 1.

Eqs. (1.2) such a rotation corresponds to an increase of amount $\Delta\varphi_2$ both in φ_1 and φ_2.

In parametric form the equations of an involute helicoid S are given by

$$\begin{aligned} x &= R_b\,[\cos\varphi_1 - (\varphi_2 - \varphi_1)\sin\varphi_1], \\ y &= R_b\,[\sin\varphi_1 + (\varphi_2 - \varphi_1)\cos\varphi_1], \\ z &= C\,[\varphi_2 - \varphi_{20}], \end{aligned} \tag{1.3}$$

where φ_{20} is a constant. Indeed a constant value of z corresponds to a value of φ_2, and the increments $\Delta\varphi_2 = \varphi_2 - \varphi_{20}$, $\Delta\varphi_1 = \Delta\varphi_2$, $\Delta z = C\Delta\varphi_2$ correspond to a rotation $\Delta\theta$ in $\Delta\varphi_2$ and an axial translation $C\Delta\theta$ of the involute $\varphi_2 = \varphi_{20}$ in the plane $z = 0$.

In applying the above to the problem of transforming a rotation about one axis A into a rotation about a skew axis A', it will be convenient to set up two fixed systems of coordinates with origins at O, O', the feet of the common perpendicular between A and A': the x, y, z system with the z-axis along the axis A and the x-axis along OO'; and the x', y', z' system with the z'-axis along A' and the x'-axis is along $O'O$. The

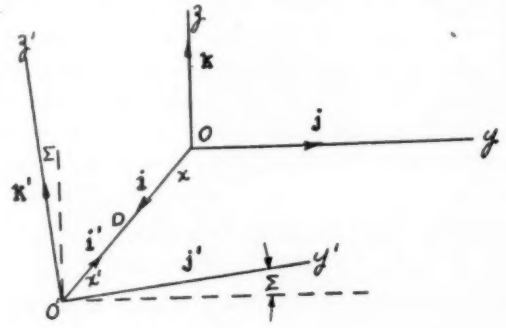

FIG. 2.

x'axis coincides with the x-axis but is reversed in direction and starts from a different origin. The relations between these coordinate systems are shown in Fig. 2, where a right-hand system is used for x, y, z, a left-hand one for x', y', z':

$$\begin{aligned} x' &= D - x, \qquad D = OO' \\ y' &= y \cos \Sigma + z \sin \Sigma, \\ z' &= -y \sin \Sigma + z \cos \Sigma. \end{aligned} \tag{1.4}$$

Here D is the mutual distance between A and A', and Σ is the angle between A and A', to be taken positive between 0 and π. The rotations ω, ω' are positive when right-handed relative to the z-, z'-axes. Unit vectors along the two sets of axes are denoted by $\mathbf{i}, \mathbf{j}, \mathbf{k}$; $\mathbf{i}', \mathbf{j}', \mathbf{k}'$ respectively; one obtains

$$\begin{aligned} \mathbf{i}' &= -\mathbf{i}, \\ \mathbf{j}' &= \mathbf{j} \cos \Sigma + \mathbf{k} \cos \Sigma, \\ \mathbf{k}' &= -\mathbf{j} \sin \Sigma + \mathbf{k} \sin \Sigma. \end{aligned} \tag{1.5}$$

Denoting by R_b' , C' the base radius and the axial displacement per unit rotation of θ' about the z'-axis of the involute helical surface S', one obtains for its equations

$$\begin{aligned} x' &= R_b' [\cos \varphi_1' - (\varphi_2' - \varphi_1') \sin \varphi_1'], \\ -y' &= R_b' [\sin \varphi_1' + (\varphi_2' - \varphi_1') \cos \varphi_1'], \\ z' &= C' [\varphi_2' - \varphi_{20}']. \end{aligned} \tag{1.6}$$

Of course, both equations (1.3) and (1.6) represent *fixed positions* of the respective helical surfaces S, S'.

For the sake of definiteness, it will be assumed that

$$| C/R_b | \geq | C'/R_b' |, \quad (1.7) \qquad R_b \geq R_b' . \quad (1.8)$$

The base radii R_b , R_b' will always be assumed to be positive; on the other hand, either C or C' may be positive or negative corresponding to right-handed or left-handed helical surfaces S, S'.

In Sec. 3 we shall prove the following theorem.

THEOREM A. *If the inequalities*

$$\Sigma < | \lambda_b | + | \lambda_b' |, \quad (1.9) \qquad | \lambda_b' | > | (\lambda_b) - \Sigma | \quad (1.10)$$

hold, it is possible to rotate the surface S' about its axis A' till it comes in point contact with the surface S at a point P_0 . If S is then rotated with a uniform velocity ω about its axis A, and if S' is allowed to rotate about its axis A' so as to maintain contact with S, then S' will rotate with a uniform velocity ω' about its axis A', where the constant ratio of velocities is given by

$$\frac{\omega'}{\omega} = \frac{C \cos \lambda_b}{C' \cos \lambda_b'} = \frac{R_b \sin \lambda_b}{R_b' \sin \lambda_b'}, \tag{1.11}$$

$$\tan \lambda_b = C/R_b , \tan \lambda_b' = C'/R_b' , [| \lambda_b | \geq | \lambda_b' | \text{ on account of (1.7)}], \tag{1.12}$$

and is independent both of the distance D between the axes and of their inclination. As the

rotation proceeds, the contact point P of S, S' describes a straight line l in space with constant linear velocity

$$v_n = \dot{C}\omega \cos \lambda_b = C'\omega' \cos \lambda_b' ; \tag{1.13}$$

this line l is the common normal to both surfaces at the initial contact point P_0, and remains normal to S, S' at P as the rotations progress.

For the limiting cases of the inequalities (1.9), (1.10) the following result is proved in Sec. 3.

THEOREM B. *Let one of the relations*

$$\Sigma = |\lambda_b| + |\lambda_b'|, \quad (1.14) \qquad \Sigma = |\lambda_b| - |\lambda_b'| \quad (1.15)$$

hold, as well as one of the relations

$$D = R_b + R_b', \quad (1.16) \qquad D = R_b - R_b'. \quad (1.17)$$

Then S' can be rotated about A' till it comes in contact with S, the contact being along a straight line g_0. If S is then rotated about A with a uniform velocity ω and S' is rotated about A' so as to maintain contact with S, then S' will rotate with a uniform velocity ω' given by (1.11). *Contact between S, S' is always along a straight line g, parallel to g_0; g moves normally to itself, in the direction of a straight line l normal to both surfaces S, S' at any point P_0 of their initial line of contact g_0, with constant velocity v_n given by* (1.13); *during this motion g describes a plane parallel to A, A' (and containing g_0 and l).*

A further limiting case of Theorem A is obtained by letting the axes A and A' become parallel. In this case the following result is obtained.

THEOREM C. *If*

$$\Sigma = 0, \qquad \lambda_b' = \pm\lambda_b, \qquad D > R_b + R_b', \tag{1.18}$$

the same conclusions hold as in Theorem B.

Two proofs of the above theorems are given in Secs. 3, 4. The proof of Sec. 4 is based upon the following theorem.

THEOREM D. *The Family $\mathfrak{F}$ of involute helicoids S with the same base cylinder and the same lead angle λ_b of the base helix forms a family of* parallel *or* normal *surfaces, that is a family of surfaces possessing the same normals, any two surfaces intersecting each normal in two points at a constant distance apart. These ∞^2 normals are all tangent to the base cylinder and make an angle λ_b with its axis. The members S of $\mathfrak{F}$ can be obtained by rotating S_0, any one member of $\mathfrak{F}$, about the cylinder axis, a rotation $\Delta\theta$ corresponding to a displacement*

$$C\Delta\theta \cos \lambda_b \tag{1.19}$$

of the intersection points along the normals.

In Sec. 3 the following theorems are also proved.

THEOREMS A', B', C'. *Under the same conditions as in Theorems A, B, C, after initial contact between S, S' has been secured, let S be simultaneously rotated about its axis A with a constant velocity $\bar{\omega}$ and translated in a direction parallel to A with a constant velocity V,*

and at the same time let S' undergo a simultaneous translation V' in direction of its axis A' and a proper rotation about A' to maintain contact with S. Then the latter rotation is a uniform rotation $\overline{\omega'}$ given by the equations

$$\omega = \bar{\omega} - V/C, \quad (1.20) \qquad \omega' = \overline{\omega'} - V'/C', \quad (1.21)$$

where ω, ω' satisfy (1.11). *From these equations $\overline{\omega'}$ can be determined in terms of $\bar{\omega}$, V, V'. As the motion proceeds the contact is of the same nature as in the respective Theorems A, B, C for the case $V = V' = 0$, and rotations ω, ω'.*

Another limiting case of interest is obtained by allowing the base radius of one of the two contacting surfaces, say S' to become infinite, while keeping its normal at a fixed point of contact with the other surface. In the limit the surface S' flattens out into a plane, the rotation of S' about its axis degenerates into a translation of the plane S' and one obtains the following theorem.

THEOREM E. *Let π be a plane tangent to S at a point P_0. Let S rotate about its axis with a uniform velocity ω, and let π, while remaining parallel to itself, be translated uniformly, normally to itself, with a velocity v_n*

$$v_n = C\omega \cos \lambda_b , \quad (1.22)$$

then π will remain tangent to S, contact occurring at any time along a straight line g which moves normally to π (and to itself) with a velocity v_n given by (1.22).

The proof of Theorem E is mentioned in Sec. 4. Its physical applications relate to mating of involute helical gears with plane racks.

Applications of the above theorems to physical gears are discussed in Sec. 5. These involve limitations due to finite axial and radial extent of the tooth surfaces, simultaneous contact between corresponding surfaces of several teeth, as well as contact to both sides of the same teeth, and avoidance of interference. The results of Sec. 5 are directly helpful in answering practical questions arising in manufacture of gears and described in the introductory section.

2. Geometry of involute helicoids. Before proceeding with the proof of the results just stated, we shall obtain several useful properties of involute helicoids.

It will be noticed that in (1.3) the coordinates are linear in φ_2. The curves $\varphi_1 =$ constant on S thus reduce to straight lines, and we conclude that an involute helicoid is a *ruled surface.*

The vectors

$$\frac{\partial \mathbf{r}}{\partial \varphi_1} = R_b (\varphi_2 - \varphi_1)(\mathbf{i} \cos \varphi_1 - \mathbf{j} \sin \varphi_1), \quad (2.1)$$

$$\frac{\partial \mathbf{r}}{\partial \varphi_2} = - \mathbf{i} R_b \sin \varphi_1 + \mathbf{j} R_b \cos \varphi_1 + \mathbf{k} C \quad (2.2)$$

where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ is the position vector of S, are tangent to the parametric (φ_1 , φ_2)-net on S; the former vector is tangent to the involutes $\varphi_2 =$ constant, $z =$ constant; the latter to the generators $\varphi_1 =$ constant. The latter vector, being dependent on φ_1 only, is constant along the generator g in question and gives the direction of g. Reduction of (2.1), (2.2) to unit vectors leads to

$$\mathbf{u} = \mathbf{i} \cos \varphi_1 - \mathbf{j} \sin \varphi_1 , \tag{2.3}$$

$$\mathbf{t} = -\mathbf{i} \sin \varphi_1 \cos \lambda_b + \mathbf{j} \cos \varphi_1 \cos \lambda_b + \mathbf{k} \sin \lambda_b , \tag{2.4}$$

where λ_b is given by (1.11).

It will be noted that

$$\mathbf{u} \cdot \mathbf{t} = 0 \tag{2.5}$$

hence the generators g are normal to the involutes z = constant.

Since the first two equations (1.3) are identical with (1.2), the projections of g on $z = 0$ are identical with the straight lines which are tangent to the base circle $r = R_b$. Hence, the generators of S are tangent to the "base cylinder" $r = R_b$. The point of contact is obtained by putting $\varphi_2 = \varphi_1$, whereupon (1.3) yields

$$x = R_b \cos \varphi_1 , \qquad y = R_b \sin \varphi_1 , \qquad z = C(\varphi_1 - \varphi_{20}). \tag{2.6}$$

This is a helix lying on the base cylinder $r = R_b$ and belonging to the screw motion of the helicoid. Its "lead angle", that is the angle it makes with the planes z = constant, is equal to arc tan C/R_b that is to λ_b. Taking differentials of (2.6) one obtains

$$d\mathbf{r} = R_b \left[-\mathbf{i} \sin \varphi_1 + \mathbf{j} \cos \varphi_1 + \mathbf{k} C/R_b\right] d\varphi_2 \tag{2.7}$$

for the tangent vector element to the helix 2(6); comparison of (2.7), (2.4) shows that both vectors are parallel, so that g is tangent to (2.6). Hence, all the generators are tangent to the helix (2.6); the latter is known as the "base helix." Thus S is a developable surface consisting of the tangents to the base helix. Therefore, S has zero (Gaussian) curvature. One of the directions of principal curvature on S corresponds to that of the generator g, with zero as the corresponding value of the principal curvature, the other principal curvature direction corresponds to the involutes sections z = constant. The generators and the involutes form the lines of curvature on S.

The normal unit vector to S is given by the cross-product $\mathbf{t} \times \mathbf{u}$:

$$\mathbf{n} = \mathbf{t} \times \mathbf{u} = \mathbf{i} \sin \varphi_1 \sin \lambda_b - \mathbf{j} \cos \varphi_1 \sin \lambda_b + \mathbf{k} \cos \lambda_b . \tag{2.8}$$

Since $\mathbf{n}$ is independent of φ_2, it follows that along every point of a generator g the normal to S has the same direction—a characteristic property of developable ruled surfaces.

Since $\mathbf{t}$ is also the unit tangent vector to the base helix, its derivative

$$\frac{d\mathbf{t}}{d\varphi_1} = -(\mathbf{i} \cos \varphi_1 + \mathbf{j} \sin \varphi_1) \cos \lambda_b \tag{2.9}$$

is readily seen to yield the direction of the principal normal of the base helix. Since $\mathbf{n}$ is normal both to $\mathbf{t}$ and to the principal (2.9), it follows that $\mathbf{n}$ is parallel to the binormal to the base helix at its point of contact with g.*

From (2.8) it will be noted that

$$\mathbf{n} \cdot \mathbf{k} = \cos \lambda_b ; \tag{2.10}$$

hence the angle between $\mathbf{n}$ and the direction of the axis A is λ_b. Thus, if from a fixed point F the various unit normals $\mathbf{n}$ to S be drawn, they will form a cone C_n of semi-

*This property, too, is characteristic of developable ruled surfaces, the binormal to the base curve C being normal to each osculating plane, and therefore normal to the ruled surface; the latter consists of tangents to C which is the envelope of the osculating planes.

angle λ_b with axis parallel to A. Similarly the unit vectors $\mathbf{t}$ giving the directions of the generators, if drawn from F, will outline a similar cone C_g, of semi-angle $\pi/2 - \lambda_b$. It will be noted that the vectors $\mathbf{k}$, $\mathbf{t}$, $\mathbf{n}$ are coplanar; hence, corresponding vectors $\mathbf{n}$, $\mathbf{t}$ on C_n, C_g are mutually perpendicular and lie in the same plane through the axis. Figure 3 shows these relations on a unit sphere.

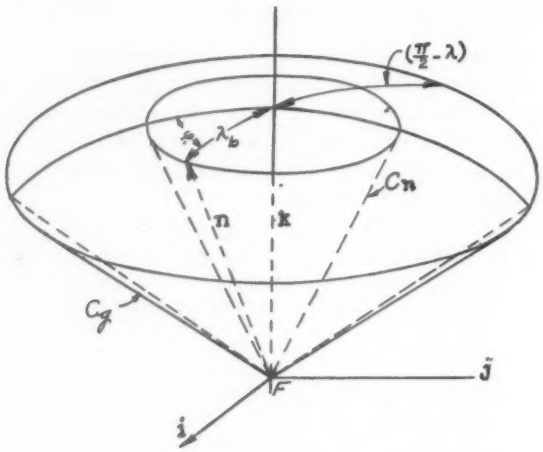

FIG. 3.

3. Contact between two involute helicoids. Contact between the two surfaces S, S' can only occur at a point at which the normal to the one surface is also normal to the other. As shown above, the unit normals $\mathbf{n}$ to S, when drawn from a fixed point F, form a cone C_n of semi-angle λ_b with axis parallel to the axis A. Likewise the unit normals to S',

$$\mathbf{n}' = \mathbf{i}' \sin \varphi_1' \sin \lambda_b' + \mathbf{j}' \cos \varphi_1' \sin \lambda_b' + \mathbf{k}' \cos \lambda_b' , \tag{3.1}$$

when drawn from the same point F, will form a cone C_n' of semi-angle λ_b' with axis parallel to A' which is parallel to

$$\mathbf{k}' = \mathbf{i} \cos \Sigma - \mathbf{j} \sin \Sigma, \tag{3.2}$$

where λ_b' is the lead angle of the generating base helix of S'. If C_n, C_n' intersect, each intersection of the two cones furnishes a possible direction for the common normal, and contact between S, S' can only occur at points at which both surfaces are normal to that direction. Figure 4 shows the two cones C_n, C_n' and their intersections FE, FE.

The rotation of the surface S about A leaves the cone C_n in Fig. 4 unchanged as a whole, though the normals rotate along the cone; similarly for rotation of S' about A' and the cone C_n'. The direction of the common normal to both mating surfaces at the point of contact, if any, thus remains invariant with time.

Several cases present themselves in regard to the cones C_n, C_n'.

I. The cones C_n, C_n' lie outside each other.

II. One of the cones C_n, C_n' lies inside the other one; on account of (1.7) C_n' will lie inside C_n.

III. The cones are tangent externally.

IV. The cones are tangent internally.

V. The cones intersect in two distinct lines.

VI. The cones coincide.

In Case I the inequality (1.9) fails to be satisfied; in Case II the inequality (1.15) holds; finally, in Case V, (1.9), (1.10) are both satisfied. Only in Cases III, IV, V, VI

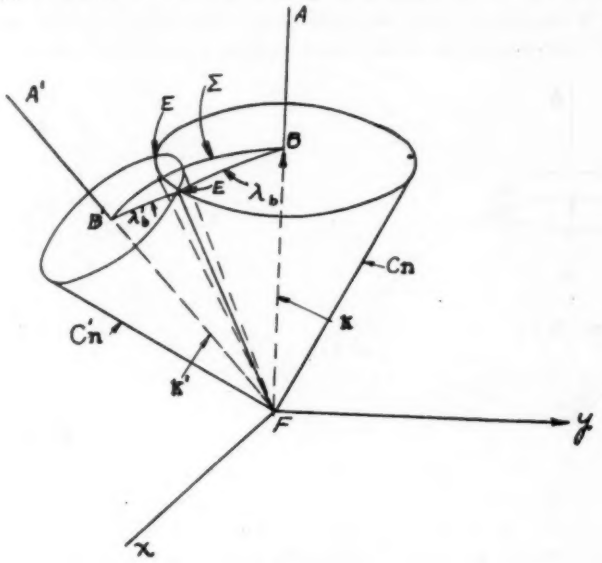

FIG. 4.

is contact between S, S' possible; III, IV lead to case discussed in Theorem B, Case V leads to the case covered in Theorem A, while VI leads to Theorem C.

The nature of contact between S and S' is exhibited in

TABLE I

Case	Theorem	Contacting Generators	Type of Contact
I		None	No contact
II		None	No contact
III	B	Parallel	Line contact
IV	B	Parallel	Line contact
V	A	Not parallel	Point contact
VI	C	Parallel	Line contact

Indeed, in Cases I, II, no common normal exists over S, S' — hence no contact between them is possible. In Case V, the common normal $\mathbf{n}$ (FE on Fig. 4) yields possible contacting generators g on S and g' on S' as the locus of points where the surface elements of S, S' are parallel. As indicated at the end of Sec. 2, the vector $\mathbf{t}$ giving the direct

of g, is coplanar with **n**, **k** and 90° away from **n**. It therefore differs from **t'**, the unit vector along g', which vector is coplanar with **n**, **k'** and 90° away from **n**. Hence, if contact can be obtained between g and g', this contact is at a point only. In Cases III, IV, as shown on Fig. 5, the vectors **n**, **k**, **k'**, are coplanar and **t**, **t'** are coincident, leading to parallel (possible) contacting generators. If contact can be secured in these cases, it will be along a line. The same statement applies in Case VI, though now the two cones C_n , C_n' of Fig. 4 degenerate into the same cone; whatever common normal one picks along this cone, one obtains the same conclusion as in Cases III, IV.

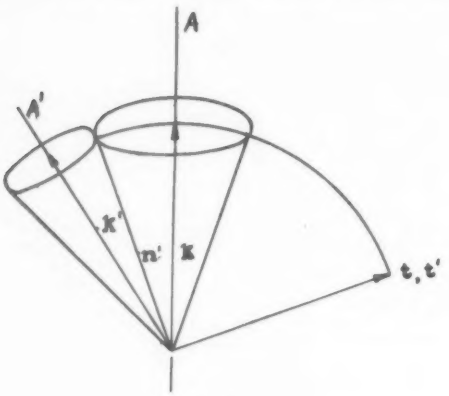

FIG. 5a.

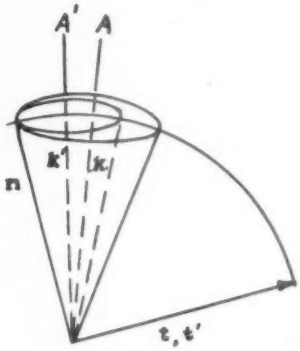

FIG. 5b.

So far it has been shown that *if* contact occurs it is of the type shown in Table I. It will now be shown that under the conditions stated in Theorems A, B, C contact between S and S' can always be secured by a proper rotation of S'.

From (1) follows that a rotation of a helical surface about its axis is equivalent to a proper translation parallel to that axis. Thus, a positive rotation of S about the z-axis through an angle $\Delta\theta$ is equivalent to a translation of S in the direction of (positive) z of amount

$$\Delta z = -C\Delta\theta. \tag{3.3}$$

Similarly, a rotation of S' about A' of angle $A\theta'$ is equivalent to a translation in the direction of positive z' of amount

$$\Delta z' = -C'\Delta\theta'. \tag{3.4}$$

Likewise, a continuous rotation of S about A, of angular velocity ω is equivalent to a translation in direction of positive z of magnitude

$$V = -C\omega, \tag{3.5}$$

and similarly a continuous rotation ω' of S' about A' is equivalent to a translation

$$V' = -C'\omega' \tag{3.6}$$

in the directions of positive z'.

Consider Case V first. Suppose that g_0 , g_0' , the initial possible contacting generators of S, S', fail to intersect. Since, as shown above, g_0' is not parallel to g_0 , and since neither

g_0 nor g_0' is parallel to A'—this follows readily from Figs. 3, 4 and the equations of Sec. 2—it is always possible to translate S' and g_0' along with it till it intersects g_0 at a point P_0. Thus in Case V, it is possible to secure contact between S' and the initial position of S by properly translating S' about A', and hence also by rotating it through a proper angle about A'. Similar considerations show that contact can be maintained as rotation of S proceeds.

The directions of $\mathbf{n}$ and hence of g, g' remain invariant with time as S, S' rotate. Figure 6 shows the contacting generators g for S in the positions at which they make

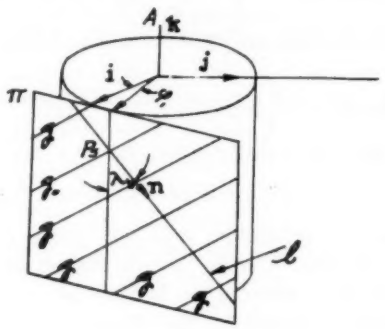

FIG. 6.

contact with corresponding generators g' or S'. If the position of the contacting generator g_0 of S is given at the time $t = 0$, the subsequent positions of $\mathbf{g}$ may be obtained by displacing it in the direction of the axis A with velocity (3.5). A similar statement applies to the contacting generator g' of S' where the displacement is in the direction of z'-axis with velocity (3.6).

As shown in Sec. 2, the generators of S are tangent to the base cylinder $r = R_b$. Therefore, as S rotates about its axis, the locus described by the contacting generator g is a plane π, through the initial position g_0 of g, parallel to A, and tangent to $r = R_b$. The plane π, in addition to being the locus of g, also contains the normals $\mathbf{n}$ to S at the points of contact P, since as shown on Fig. 3 $\mathbf{n}$, $\mathbf{k}$, $\mathbf{t}$ are coplanar. Likewise, the contacting generator g' of the surface S', as the latter rotates, describes a plane π' tangent to the base cylinder of S' and containing the directions of A', of g', and the normal $\mathbf{n}$. The line of intersection l of π, π' forms the locus of the contact point P. Since each of the two planes π, π' is parallel to $\mathbf{n}$, their line of intersection l is itself parallel to $\mathbf{n}$.

The line l is tangent to the base cylinder of S at P_S and normal to its base helix there. Being parallel to $\mathbf{n}$, l is normal to S at the contact point as S undergoes its rotation or equivalent displacement. The angle between l and the z-axis is the same as between $\mathbf{n}$ and $\mathbf{k}$, hence equal to λ_b. Hence, the translation (3.5) of the contacting generator g leads to a motion of the contact point P along l with velocity

$$V \cos \lambda_b = -C\omega \cos \lambda_b . \tag{3.7}$$

Similarly, the translation (3.6) of the contacting generator of S' leads to the motion of the contact point along l with velocity

$$V' \cos \lambda_b' = -C'\omega' \cos \lambda_b' . \tag{3.8}$$

Equating the two velocities and denoting their common value by v_n , one arrives at (1.13) and (1.11).

This essentially completes the proof of Theorem A.

The cone C_n of unit normals $\mathbf{n}$ to S is given by (2.8) with variable φ_1 . The intersections of the two cones C_n , C_n' are obtained by imposing the condition

$$\mathbf{n}\cdot\mathbf{k}' = \cos\varphi_1 \sin\lambda_b \sin\Sigma + \cos\lambda_b \cos\Sigma = \cos\lambda_b' . \tag{3.9}$$

Solving for φ_1 and substituting in 2(8) yields the two possible common normals to S and S'.

Equating $\mathbf{n}'$ in 3(1) to $\mathbf{n}$ in 2(8) and expressing $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ in terms of $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ one is led to the following relations:

$$\begin{aligned} \sin\varphi_1' \sin\lambda_b' &= -\sin\varphi_1 \sin\lambda_b , \\ \cos\varphi_1' \sin\lambda_b' &= -\cos\varphi_1 \sin\lambda_b \cos\Sigma + \cos\lambda_b \sin\Sigma, \\ \cos\lambda_b' &= \cos\varphi_1 \sin\lambda_b \sin\Sigma + \cos\lambda_b \cos\Sigma. \end{aligned} \tag{3.10}$$

The third of these relations is the same as (3.9), and will be recognized on the spherical triangle BEB' of Fig. 4 and Fig. 7 as the law of cosines for side $\lambda_b = EB$. The first equation (3.10) determines φ_1' in terms of φ_1 , λ_b , λ_b' , and forms the law of sines for the same triangle for the angles B, B' and the sides opposite.

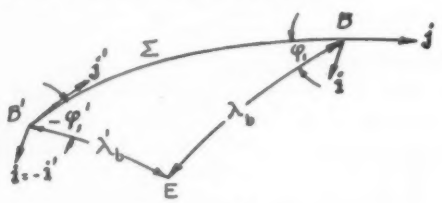

FIG. 7.

Substitution of the value of φ_1 obtained from (3.9) in (2.6) yields the point on the base helix of S in the plane π. Since this plane is tangent to $r = R_b$ at this point, its equation is

$$\pi: \quad x\cos\varphi_1 + y\sin\varphi_1 = R_b . \tag{3.11}$$

Both the contacting generator g and the line of contact l lie in π. The line l and g' lie in the plane π' whose equation is

$$\pi': \quad x'\cos\varphi_1' - y'\sin\varphi_1' = R_b' , \tag{3.12}$$

where φ_1' is determined by means of (3.10). Utilizing (1.4) one changes (3.12) into

$$x\cos\varphi_1' + y\cos\Sigma\sin\varphi_1' + z\sin\Sigma\sin\varphi_1' = D\cos\varphi_1' - R_b' . \tag{3.13}$$

Turn now to Cases III, IV. Referring to Figures 5, we now have for the common normal

$$\mathbf{n} = -\mathbf{j}\sin\lambda_b + \mathbf{k}\cos\lambda_b \tag{3.14}$$

corresponding to $\varphi_1 = 0$ in (2.8), and $\varphi_1' = 0$ in (3.1). All the vectors $\mathbf{k}$, $\mathbf{k}'$, $\mathbf{n}$, $\mathbf{t}_2 = \mathbf{t}_2'$

lie in the plane normal to **i**. Hence both planes π and π', the locus of the possible contacting generators, are parallel; in fact their equations (3.11), (3.13) now reduce to

$$\pi: \quad x = R_b\,, \quad (3.15) \qquad \pi': \quad x = D \pm R_b'\,. \quad (3.16)$$

Clearly, unless π, π' coincide, no amount of translation of S' along A', or of equivalent rotation about A', will bring g' into contact with g. On the other hand, if π and π' do coincide, thus leading to (1.16) or (1.17), a proper translation of S' along A' will bring the possible contacting generator g_0' into coincidence with g_0 . From this point on, the rest of Theorem B is proved similarly to Theorem A.

As regards Case VI and parallel axes A, A', one now finds that with coincident cones C_n , C_n' the degenerate Fig. 4 or equation (3.9) fails to pick common normals or possible contacting generators. Whatever normal **n** is assumed it leads to $\mathbf{t}_2 = \mathbf{t}_2'$, to parallel possible contacting generators g, g', and to parallel planes π, π'. For contact to occur, the planes π, π' must reduce to the common tangent planes of the two base cylinders:

$$\frac{R_b}{\cos\varphi_1} \pm \frac{R_b'}{\cos\varphi_1} = D. \qquad (3.17)$$

With φ_1 determined from (3.17) the rest of the proof of Theorem C follows along the same lines as above. The last inequality (1.18) prevents the base cylinder $r = R_b$ from enclosing the other one, thus rendering common tangent planes non-existent. Contact occurs now not at a point but along a line g, and as rotation proceeds g describes the plane π which is identical with π' and is given by (3.17).

Theorems A', B', C' are proved by utilizing the equivalence of rotations and translations.

4. Alternative proof based on theorem D. To establish Theorem D, it will now be shown that the normal l to the surface S at a point P_0 on it is also normal to each involute helicoid of the family $\mathfrak{F}$, given by (3.1) for other values of the constant φ_{20} , at its point of intersection with that surface. The members of the family $\mathfrak{F}$ can be obtained by translating S along the z-axis through a variable distance Δz, or by rotating S about the z-axis through a variable angle $\Delta\theta$. They all have the same base cylinder and the same base lead angle λ_b , but different base helices.

Let P_0 be a point on S obtained by putting $\varphi_1 = \varphi_1^0$, $\varphi_2 = \varphi_2^0$ in (1.3):

$$\overrightarrow{OP}_0 = \mathbf{r}_0 = \begin{cases} R_b\,[\cos\varphi_1^0 - (\varphi_2^0 - \varphi_1^0)\sin\varphi_1^0], \\ R_b\,[\sin\varphi_1^0 + (\varphi_2^0 - \varphi_2^0)\cos\varphi_1^0], \\ C[\varphi_2^0 - \varphi_{20}]. \end{cases} \qquad (4.1)$$

The unit normal along the line l, normal to S at P_0 is obtained from (2.8):

$$\mathbf{n} = (\sin\varphi_1^0 \sin\lambda_b\,, \qquad -\cos\varphi_1^0 \sin\lambda_b\,, \qquad \cos\lambda_b). \qquad (4.2)$$

Now consider the point P on l a distance s from P_0 :

$$\overrightarrow{0P} = \mathbf{r}_0 + s\mathbf{n}. \qquad (4.3)$$

The coordinates of P can also be obtained from (1.3) by increasing φ_1 , φ_2 , φ_{20} from their values at P_0 by the amounts

$$\Delta\varphi_1 = 0, \qquad \Delta\varphi_2 = -(s/R_b)\sin\lambda_b\,, \qquad \Delta\varphi_{20} = -(s/C)\sec\lambda_b \tag{4.4}$$

and P thus lies on a surface of the family $\mathfrak{F}$. The normal to this surface at P, moreover, agrees with (4.2) since $\varphi_1 = \varphi_1^0$. Thus l is normal to the helical involute surface passing through P.

From the above it follows that the family $\mathfrak{F}$ is a family of *parallel* surfaces admitting the same normals, each surface being a constant distance apart from every other surface. The directions of principal curvatures along such surfaces at their points of intersection with the same normal are parallel and the principal radii of curvatures at these points differ by the constant normal distance.

The common normals l to this family $\mathfrak{F}$ of parallel surfaces are all tangent to the base cylinder and make an angle with the direction of its axis A equal to λ_b. Stated in another way, it has been shown that *the surfaces perpendicular to the lines which are tangent to the cylinder $r = R_b$ and make a fixed angle λ_b with its axis form a family of involute helicoids with their base helices on $r = R_b$ and of lead angle λ_b.*

Suppose next that S rotates about its axis A with uniform velocity ω. It will be shown that its intersection with the normal l at the point P_0 moves with uniform velocity

$$v = -C\omega\cos\lambda_b\,. \tag{4.5}$$

Indeed, putting

$$s = vt = -C\omega t\cos\lambda_b = -R_b\omega t\sin\lambda_b \tag{4.6}$$

in (4.3), (4.4) leads to the above point P on the normal; it is readily verified that the same point is obtained if in (1.3) we first put

$$\varphi_1 = \varphi_1^0 - \omega t, \tag{4.7}$$

$$\varphi_2 = \varphi_2^0 - \omega t\cos^2\lambda_b \tag{4.8}$$

(while keeping φ_{20} fixed), then, corresponding to the rotation ωt, we increase φ_1, φ_2 by ωt in the first two Eqs. (1.3), thus putting

$$\varphi_1 = \varphi_1^0\,, \tag{4.9}$$

$$\varphi_2 = \varphi_2^0 + \omega t\sin^2\lambda_b \tag{4.10}$$

there, while allowing (4.8) to hold in the third equation.

There results

$$\begin{aligned} \Delta x &= -R_b\omega t\sin^2\lambda_b\sin\varphi_1^0\,, \\ \Delta y &= R_b\omega t\sin^2\lambda_b\cos\varphi_1^0\,, \\ \Delta z &= -C\omega t\cos^2\lambda_b \end{aligned} \tag{4.11}$$

and these agree with $s\mathbf{n}$ for s given by (4.6).

Thus a uniform rotation of S leads to a point of intersection with the normal l which moves with a uniform velocity along l.

By applying this result to two involute helical surfaces S, S' initially in contact at a point or a line, Theorems A, B, C follow.

Theorem E follows readily from D by considering a rotation of S, its intersection with a fixed normal line l and the tangent plane to S which is normal to l at that point of intersection.

5. Application to gears. In applying the results of the preceding sections to two mating gears G, G', certain limitations occur due to the finite dimensions of the gears, to non-interference of mating teeth, and to the integral number of teeth on each gear.

Let the gears G, G' possess an integer number of teeth, N, N' respectively, and suppose that the mating teeth are properly shaped so that a uniform rotation of G is transformed into a uniform rotation ω' of G'. After a time T corresponding to N' complete rotations of G, the number of teeth of G that have gone through mesh is NN'. During the same time T a like number NN' of teeth of G' has gone through mesh; hence G' has gone through N complete rotations. Thus, the number of complete rotations of G, G' in the time T varies inversely as the number of teeth, and hence

$$\frac{\omega'}{\omega} = \frac{N}{N'}. \tag{5.1}$$

For involute helicoid tooth shape, combining (5.1) with (1.11), one obtains

$$\frac{\omega'}{\omega} = \frac{N}{N'} = \frac{C \cos \lambda_b}{C' \cos \lambda_b'} = \frac{R_b \sin \lambda_b}{R_b' \sin \lambda_b'}. \tag{5.2}$$

This is a restriction on the design brought about by the integral numbers of teeth.

In actual gears the radii r, r' of the involute surfaces are limited by the dedendum and addendum radii, and continuity of rotations is assured by having *several teeth* in mesh part or all of the time. This is obtained by having the angle of contact exceed $2\pi/N$. Since, in accordance with Theorem A, the point of contact moves with velocity (1.13) along the line of contact l, this condition is assured by having a segment on the line of contact greater than

$$\frac{2\pi}{N} C \cos \lambda_b = d \tag{5.3}$$

lie both within the limits of r, r' corresponding to the mating teeth and within the z, z' limits corresponding to the face width. For involute helicoid gears satisfying (5.2), *no new* conditions are required for assuring correct *simultaneous contact* between corresponding surfaces of several teeth. Two involute surfaces S spaced $2\pi/N$ radians apart, according to Sec. 4, are at a constant (normal) distance apart, equal to d. Hence, condition (5.2) assures one that for two adjacent tooth surfaces on G' the distance d' agrees with the corresponding d for the mating surfaces of S; hence, simultaneous contact takes place without interference.

If the gears have no backlash, contact on *opposite sides* of various teeth occurs, and the question of interference between the mating action of opposite sides arises. These opposite sides on S are characterized by the respective inequalities

$$\varphi_2 > \varphi_1, \tag{5.4}$$

$$\varphi_2 < \varphi_1 \tag{5.5}$$

in Eqs. (1.3) (see Fig. 8). This is evident on Fig. 1 where the cusp of the involute is seen to correspond to

$$\varphi_1 = \varphi_2. \tag{5.6}$$

The theorems and proofs of Sec. 1–4 now have to be reviewed

from point of view of the inequalities (5.4), (5.5) and the corresponding inequalities

$$\varphi_2' > \varphi_1' , \tag{5.7}$$

$$\varphi_2' < \varphi_1' \tag{5.8}$$

in Eqs. (1.6) for the surfaces of the teeth of S'.

As stated in Sec. 1, only if one of the inequalities (5.4), (5.5) holds will the orthogonal net of involutes φ_2 = const., and the tangents to the base circle, φ_1 = const., cover

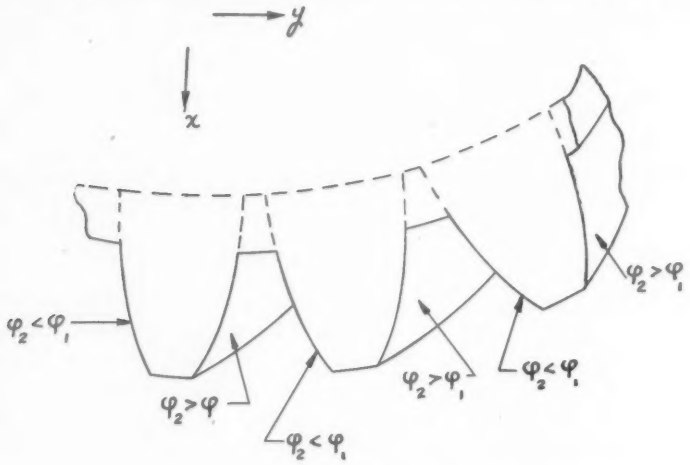

FIG. 8.

the region outside the base circle so that through each point there passes one and only one involute, and one and only one tangent line. Likewise, only if *one* of the inequalities (5.4), (5.5) holds does there result a family of involute helicoids $\mathfrak{F}$ such that through each point outside the base cylinder there passes one and only one member of this family. The two involute helicoids through a point satisfying (5.4), (5.5) respectively have distinct normals. In particular, the two intersections of the cones C_n , C_n' on Fig. 4 correspond respectively to upper and lower faces of gear teeth of G. These distinct normals are obtained from (3.9) by choosing the positive and the negative values of φ_1 . Each one of the normals leads to a distinct line l and distinct planes π, π'. However, on account of the inequalities (5.4), (5.5), (5.7), (5.8), contact along each line l is restricted to a portion lying to the proper side of each of the two base cylinders. (In practice this segment is even further restricted by the limits of r, r'; z, z' of the teeth of G, G'.)

As an example, consider the technically important case where one of the two contact lines l cuts the x-axis, that is where the line of contact intersects the common perpendicular of the axes A; A'. Putting $y = 0$, $z = 0$ in (3.11), (3.13), one obtains for the resulting contact point

$$x = R_b/\cos \varphi_1 , \qquad y = 0, \qquad z = 0 \tag{5.9}$$

as well as

$$x' = (D - x) = R_b'/\cos \varphi_1' , \qquad y' = 0, \qquad z' = 0, \tag{5.10}$$

whence

$$D = \frac{R_b}{\cos \varphi_1} + \frac{R_b'}{\cos \varphi_1'} . \tag{5.11}$$

The last restriction on the distance D furnishes the condition for the intersection of l and the x-axis. It will be noted now that if (5.11) is satisfied, it will still be valid if the signs of φ_1 , φ_1' are changed. Hence, *both* lines of contact l will intersect the x-axis if one of them does.

Substituting (5.9) in (1.3) one obtains

$$\varphi_2 = \varphi_1 - \tan \varphi_1 . \tag{5.12}$$

Hence the sign of φ_2 is opposite to that of φ_1 , so that if $\varphi_1 > 0$ Eq. (5.5) holds, while for $\varphi_1 < 0$ Eq. (5.4) applies. If we suppose further that G, G' are right handed and $\Sigma < \pi/2$ it will be seen that the line l for which $\varphi_1 > 0$ corresponds to contact between an upper face of G and a lower face of G'. This checks with the first Eq. (3.10) according to which φ_1' is now negative. The line l with $\varphi_1 < 0$ yields the contact between a lower face of G and an upper face of G'. Both lines l pass through the point (5.9), which is now known as the "common pitch point" of G, G'—while their directions are given by (2.8) for the proper value of φ_1 .

The function $\tan \varphi - \varphi$ is sometimes denoted by "inv φ" (pronounce the "involute function of φ"), on account of its frequent occurrence in involute theory.

It is customary to introduce the "pitch radii"

$$R_p = R_b/\cos \varphi_1 , \quad (5.13) \qquad R_p' = R_b'/\cos \varphi_1' , \quad (5.14)$$

as the distances from the common pitch point to the axes A, A'. The helices of S, S' corresponding to the pitch radii are known as the "pitch lines." The lead angle of the pitch line of S is denoted by λ_p :

$$\tan \lambda_p = C/R_p = (C/R_b)(R_b/R_p) = \tan \lambda_b \cos \varphi_1 \tag{5.15}$$

and λ_p' is introduced similarly.

The normal projection of (either) common normal $\mathbf{n}$ on a plane $x =$ constant, it will be noted from (2.8), makes an angle

$$\tan^{-1} (\sec \varphi_1 \tan \lambda_b) = \lambda_p \tag{5.16}$$

with the z-axis; it is thus normal to the pitch line of S; likewise, it is normal to the pitch line of S' through the same point. There is thus contact between the pitch lines of S, S' through the pitch point (5.9) at the instant of contact of S, S' at that point. This condition leads to

$$\Sigma = \lambda_p + \lambda_p' \tag{5.17}$$

—a relation that also follows from (3.10) by expressing λ_b , λ_b' in terms of λ_p , λ_p' . The angle $| \varphi_1 |$ itself is now readily shown to be equal to the "pressure angle" at the pitch radius in the transverse section $z = 0$, that is to the angle between the radial direction and the involute through the pitch point. The pressure angles $| \varphi_1 |$, $| \varphi_1' |$, the "lead angles" λ_p , λ_p' , and the pitch radii R_p , R_p' are the quantities technically used in specifying helical gears.

For the case of pitch-point contact, the condition of simultaneous contact on both sides of the teeth of G, G' is given by

$$\frac{(\Delta\theta)_p}{2\pi/N} = 1 - \frac{(\Delta\theta')_p}{2\pi/N'} \tag{5.18}$$

where $(\Delta\theta)_p$, $(\Delta\theta')_p$ are the angular thicknesses of the teeth G, G' at their pitch radii; thus, $(\Delta\theta)_p$ is the increase in θ along $r = R_p$, $z = 0$ in passing from one side of a tooth of G to the other side. This equation is obtained by noting that since now the two lines of contact l both pass through the common pitch point (5.9), the time that it takes a tooth of G to pass over this point is equal to the time it takes the space between two adjacent teeth of G' to sweep over the same point.

To obtain a relation equivalent to (5.18) but involving angular tooth thickness at other radii than the pitch radius, the variation of tooth thickness with r is obtained from Eqs. (1.2), keeping in mind that along an involute φ_2 is constant. From these equations follow

$$r = R_b[1 + (\varphi_2 - \varphi_1)^2]^{1/2}, \tag{5.19}$$

$$\tan\theta = \tan\varphi_1 + (\varphi_2 - \varphi_1) = \varphi_2 + \text{inv}\,\varphi_1 \tag{5.20}$$

and after determining the change in φ_1 corresponding to a change in r, one may then obtain the corresponding change in θ from (5.20). Multiplication by 2 then leads to the change in tooth thickness.

To examine the case when (5.11) is not satisfied it is convenient first to shift the axis A', while keeping it parallel to itself, to a position for which (5.11) does hold, then move it back a distance

$$\Delta D = D - R_p - R_p' = D - R_b/\cos\varphi_1 - R_b'/\cos\varphi_1' \tag{5.21}$$

to its original position. Since the cones C_n , C_n' of Fig. 4 are unchanged by this displacement, the direction of $\mathbf{n}$, the values of φ_1 , φ_1' which satisfy (3.9), as well as the directions of l and the possible contacting generators g, g' are not affected, except, of course, for the relative displacement of the latter. Comparing the solutions of (3.11), (3.12) for A' and for its shifted position, it will be noted that a solution is offered by the same x, y as for the pitch-point contact case, provided that z be increased by the amount

$$\Delta z = \Delta D \cot\varphi_1'/\sin\Sigma. \tag{5.22}$$

It will be noted that corresponding to a change of sign of φ_1 there is also a change of sign of Δz, that is that the *two* contact lines l are displaced in *opposite* directions from the point (5.10) as D is changed in value.

Let

$$n_x x + n_y y + n_z z = p \tag{5.23}$$

be the equation of the common tangent plane at a point of contact of the two gears G, G', where n_x , n_y , n_z are the components of the normal to the plane, and p the distance of the plane from the origin. Suppose now that G' undergoes a displacement $\Delta D\mathbf{i}$, thus throwing the gears out of contact, and displacing the tangent plane to G' from (5.24) to

$$n_x(x - \Delta D) + n_y y + n_z z = p, \tag{5.24}$$

or, transposing, to

$$n_x x + n_y y + n_z z = p + n_z \Delta D. \tag{5.25}$$

Since φ_1 , φ_1' are not effected by the displacement $\Delta D\mathbf{i}$, the possible contacting generator g' undergoes the same displacement as G', while g remains unchanged. Now note that (5.25) may also be obtained by increasing z by the amount

$$\Delta z = \Delta D \, n_z/n_z \, ; \tag{5.26}$$

hence contact may be restored by displacing G in the direction of positive z by the amount (5.26). Substituting from (2.8) for n_x , n_z and recalling that a displacement of G in the direction of its axis A is equivalent to a proper rotation about A, it follows that

$$\frac{\Delta D}{C} \sin \varphi_1 \tan \lambda_b \tag{5.27}$$

is the rotation of G required to close up the gap due to the displacement ΔD of G, and restore contact with G'. Change of sign of φ_1 corresponding to contact on the other tooth side causes the rotation (5.27) to change in sign. Hence

$$\frac{2\Delta D}{C} \sin \varphi_1 \tan \lambda_b = \frac{2\Delta D}{R_b} \sin \varphi_1 \tag{5.28}$$

is the backlash between G, G' (measured as a rotation of G about its axis) produced by the displacement $\Delta D\mathbf{i}$ from initial zero backlash position. Again, (5.28) is the increased angular thickness of a tooth of G caused by a displacement $\Delta D\mathbf{i}$ of a generator tool (involute hob, involute shaving tool) having the shape of G'.

Similarly it is shown that the same backlash is measured by the rotation

$$\frac{2\Delta D}{C'} \sin \varphi_1' \tan \lambda_b' = \frac{2\Delta D}{R_b'} \sin \varphi_1' \tag{5.29}$$

of G' about its axis; this is the required thickening of teeth of G' to produce the same thickness tooth on G after the displacement $\Delta D\mathbf{i}$ has taken place.

These equations along with (5.18) and the general theory outlined above, suffice for the design of hobs and shaving tools for producing involute helical teeth and for investigating the effect of displacement of axes and angles, of resharpening dulled tools, etc.

As an example, suppose two external gears G, G', designed for pitch-point contact, are mounted so that their axes A, A' are at the correct distance D apart but with a wrong relative inclination, so that the angle $\overline{\Sigma}$ between A, A' is incorrect and (5.17) does not hold for $\overline{\Sigma}$. What effect does this have on the mating action of the gears?

Since Eqs. (5.3), (5.3) are not affected by the angle between the axes, correct mating action with the proper ratio of speeds ω/ω' is still possible for *either* one side of the teeth or the other, but the matter of contact on *both* sides or of backlash or interference due to insufficient space between the teeth still remains to be examined. In the spherical triangle $BB'E$ of Figs. 4, 7 the sides λ_b , λ_b' are unaffected, but with Σ now changed to $\overline{\Sigma}$ new values $\overline{\varphi}_1$, $\overline{\varphi}_1'$ for φ_1 , φ_1' result; these may be determined from the last and the first equations (3.10). For these values $\overline{\varphi}_1$, $\overline{\varphi}_1'$ and $\overline{\Sigma}$ it is possible to replace the nominal pitch radii R_p , R_p' by new pitch radii $\overline{R}_p$, $\overline{R}_p'$ determined from (5.13), (5.14) and to consider a new pitch-point contact condition with a distance $\overline{D}$ between the axes de-

termined from (5.11). Assuming a zero backlash condition, and supposing that, say, G' has the correct tooth thickness, one determines its value $(\Delta\bar{\theta}')_p$ at $\bar{R}_p'$, and applying (5.18), one obtains tooth thickness $\Delta\bar{\theta}_p$ at $\bar{R}_p$. Finally, allowing the gear G' to undergo a displacement $\Delta D\mathbf{i} = (D - \bar{D})\mathbf{i}$ till the distance between A, A' has its correct initial value D, one adds (5.28) to $\Delta\bar{\theta}_p$ to obtain the tooth thickness of G at $\bar{R}_p$ for the zero-backlash condition corresponding to D, $\bar{\Sigma}$, and this may be compared with the design tooth thickness of G. If G' corresponds to a hob or to a shaving tool of involute profile, the above analysis shows that if it is set at an incorrect angle $\bar{\Sigma}$ or at a wrong distance D, (or both), a correct involute gear shape will be produced, but possibly a wrong tooth thickness. (There is also, of course, the effect of varying depth of penetration.)

The analysis of Secs. 3, 4 for obtaining contact between mating gears can be regarded as the equivalent of the *first necessary condition* for such contact and avoidance of interference; it is analogous to the vanishing of the first derivative $f'(x)$ for conditions of obtaining a maximum of a function $f(x)$, or more precisely, to the vanishing of the two partial derivatives $\partial f/\partial x$, $\partial f/\partial y$ for a maximum of a function $f(x, y)$ of two variables x, y. In such maxima problems, in addition to the first necessary conditions, there also exist other conditions, for instance, sufficient conditions for a local maximum involving second derivatives. Similarly, in the problem of contact of two physical gears G, G' further conditions have to be examined in addition to the tangency of the tooth surfaces, to avoid interference and assure that at no time is any point inside the gear G trying to occupy a position inside G'.

A case of great technical importance is the case of contact of two external gears under the conditions of Theorem A and Case V, if

$$D > R_b + R_b', \tag{5.30}$$

and if contact is of the common pitch-point type or "near it", that is reducible to pitch-point contact by a continuous change of D and Σ. It is with this important case in mind that Fig. 2 was drawn, and the directions of axes on it were chosen, the positive directions of x, x' pointing toward the region of contact. For this case the second-order conditions are easily verified by considering the curvatures of the two involute helical surfaces at the contact points. As will be recalled from Sec. 2, an involute helical surface is a ruled surface. For each of the contacting helical surfaces at the point of contact, the lines of zero curvature are the generators g, g', and they lie in the common tangent plane. For external gears in contact under the conditions stated, the centers of curvature for the second principal direction of curvature lie to opposite sides of the common tangent plane. This insures avoidance of interference near the contact point.

Under the condition (1.16) of Theorem B it turns out that the zero curvature directions are coincident (along the coincident rulings g, g'), but that the centers of curvature for the second principal curvature for S, S' lie on the *same* side of the tangent plane. It is therefore impossible to realize the type of contact envisioned in Theorem B between two external gears G, G'. This statement applies quite aside from the interpenetration of the teeth of one gear into the base circle of the other gear near the region of the common perpendicular to their axes (the x-axis).* Even if the active face widths (the limits of z and z' for the involute teeth) for G, G' are restricted to lie sufficiently far from the x-axis to avoid this obvious interference, it still turns out that the contact considered

*To a certain extent this may be avoided by proper undercutting of the teeth.

in Theorem B cannot be realized for external gears. Any attempt to run a pair of external gears under these conditions would merely result in contact of the edge of the teeth of one against the tooth surfaces of the other. Similar statements apply to Theorem A for external gears when the inequality

$$D < R_b + R_b' \tag{5.31}$$

holds.

If (5.2), (5.30) holds, under the conditions (1.18), and if contact is of the common pitch-point type or near it the conditions considered in Theorem C *can* be realized between external gears G, G', and contact always *is* of the common pitch-point type, irrespective of the value of D.

If G' is an external gear and G is an internal one, then the contact conditions of Theorem A generally cannot be realized physically due to interference.†

Thus, in shaving internal involute gears with involute helical cutters, a condition of theoretical interference is obtained. In some cases the amount of this interference is slight enough not to cause any trouble. In other cases the shaving cutter teeth are "relieved" at the ends to avoid interference. This relieving operation, of course, tends to destroy the involute helical shape.

If G is internal and G' external, the conditions of contact of Theorems B, C may be realized in certain cases, though, of course, the condition (5.30) has to be dispensed with for Theorem C.

Even for external gears with (5.30) holding other types of contact than of the "near common pitch-point contact" type are possible. This is not evident from the proofs of Secs. 3, 5, from which it would appear that in case of Theorem A *only two* lines of contact l are possible in the mating action of two gears G, G'. Indeed, Fig. 4 and Eq. (3.10) determine two possible directions of the normal and the values of φ_1 , φ_1' in terms of λ_b , λ_b' , Σ, while Eqs. (3.11), (3.13) then fix a line l uniquely for each pair φ_1 , φ_1' . Actually, however, further contact lines l may exist due to the fact that on Fig. 4 *complete cones* and *not half-cones* C_n , C_n' should be used, that is cones proceeding *both ways* from the vertex F. Indeed, for a given φ_1 , λ_b , not only is a possible unit normal $\mathbf{n}$ given by (2.8), but also by its negative. If the signs of both $\mathbf{n}$ and $\mathbf{n}'$ are reversed, two new circles of intersection of C_n , C_n' with the unit sphere are obtained on Fig. 4. The new circle for C_n may intersect the (old) circle for C_n' in two new points, thus giving rise to two further possible directions for the common normals. To obtain these (when they exist) it is necessary to modify Eqs. (3.10) by changing the signs of their right-hand members. With the new values of φ_1 , φ_1' so obtained, Eqs. (3.11), (3.13) lead to two new contact lines l.

As a special case, if with the resulting new values of φ_1 , φ_1' , D is given by

$$D = \left| \frac{R_b}{\cos \varphi_1} - \frac{R_b'}{\cos \varphi_1'} \right| , \tag{5.32}$$

then the two contact lines l pass through a point Q on the line $00'$ lying outside the segment $00'$. The point Q is analogous to the common pitch point P which exists when

†It must be borne in mind that due to elastic yielding at contacting areas (in accordance with the Hertz theory) the purely geometric conclusions regarding what type of contact is or is not physically possible may be softened somewhat.

(5.11) holds. When (5.32) does not hold the new contact may be determined by displacing the lines l through Q by amounts (5.22) in the direction of the z-axis.

For the special case of parallel axes A, A', as pointed out above, the cones C_n , C_n' of Fig. 4 degenerate into coincident cones. Now contact is always of the pitch point type and contact lines passing through both P and Q are possible. On Fig. 9 are shown the projections of the lines l on the plane $z = 0$.

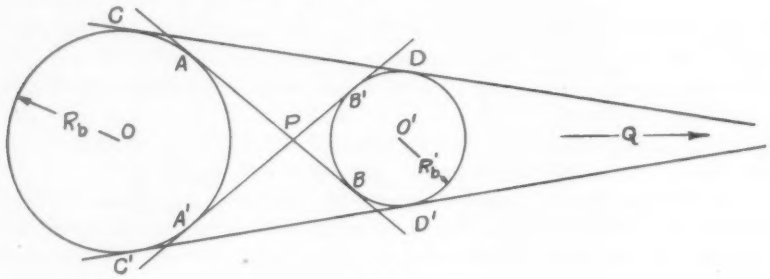

FIG. 9.

The point Q leads to a negative ratio ω/ω'; any attempt to realize contact along the tangent lines CD, $C'D'$ through it by means of external gears is necessarily confined to a portion between the points of tangency with the base circles, and therefore does not include the point Q itself as a contact point. It will be recognized that, due to interference and other difficulties, the angle of rotation will be limited.

Similarly, for helical involute gears and skew axes it is possible to have contact lines l which intersect the common perpendicular between the two axes outside OO'. While the Theorem A can be applied to this type of contact, it again does not lead to generally useful results.

THE RADIATION AND TRANSMISSION PROPERTIES OF A PAIR OF PARALLEL PLATES-II*

BY

ALBERT E. HEINS

Carnegie Institute of Technology

1. Introduction. In Part I, we discussed the excitation of a pair of semi-infinite parallel plates by an electromagnetic plane wave which had only one component of the electric field. We now reverse our discussion in order to find the effect of a parallel plate mode on the excitation. In Fig. 1 we assume that the structure has been excited for $z \ll 0$ by a mode of the form $e^{ikz} \sin (\pi x/a)$, that is a parallel plate mode traveling to the right. A reflected mode is excited at the mouth of the parallel plate region and has the form $e^{-ikz} \sin (\pi x/a)$ for $z \ll 0$. We are now interested in calculating the reflection coefficient, that is the ratio of the amplitude of the reflected parallel plate mode $e^{-ikz} \sin (\pi x/a)$ to the amplitude of the incident one, $e^{ikz} \sin (\pi x/a)$. Because of the presence of certain symmetries we shall find that this problem may be formulated as a single integral equation. We shall find that much of the mathematical technique which we developed in Part I carries over, so that we shall not have to discuss this part in great detail. The author wishes to thank Dr. J. F. Carlson and Dr. J. S. Schwinger for several stimulating discussions on this problem.

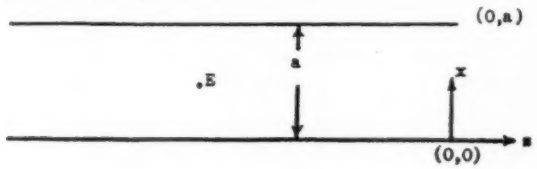

FIG. 1.

2. The formulation of the problem. The structure in Fig. 1 has been excited symmetrically about the line $x = a/2$ by an electric field of the form $E_y(x, z) = \rho_1 e^{+ikz} \sin (\pi x/a)$. As such, no even modes containing factors of the form $\exp\{z[(4m^2\pi^2/a^2) - k^2]^{1/2}\} \sin (2m\pi x/a)$, $z \ll 0$ will be excited in the parallel plate region.[1] The line $x = a/2$ is then a line of maximum electric intensity or zero tangential magnetic intensity. Figure 1 may therefore be replaced by Fig. 2. The condition for $0 \leq x \leq a/2$, $z \ll 0$

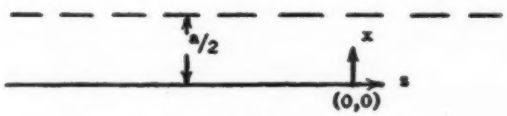

FIG. 2.

still holds. That is, $E_y(x, z)$ is asymptotic to $\sin (\pi x/a)(\rho_1 e^{ikz} + \rho_2 e^{-ikz})$. On the plane

*Received March 29, 1947. Presented to the American Mathematical Society on Dec. 27, 1946. Part I of this paper appeared in this Quarterly **5**, 157-166 (1948).

[1]It is assumed as in Part I that only the lowest mode propagates.

$x = 0, z < 0$, $E_v(x, z)$ vanishes, while on the plane $x = a/2, -\infty < z < \infty$, $\partial E_v/\partial x$ vanishes identically in x and z. Since there is no plane wave field external to the parallel plate region, $E_v(x, z)$ is at most of the form of a radiated field and is asymptotic to $e^{ikr}/r^{1/2}$ for $r \gg 0$, where $r = (x^2 + z^2)^{1/2}$.

We first divide the half plane $x < a/2, -\infty < z < \infty$ into two regions, the strip $-\infty < z < \infty, 0 \leq x \leq a/2$, and the half plane $-\infty < z < \infty, x < 0$. It is now possible to express $E_v(x, z)$ in each region in terms of an appropriate Green's function and $E_v(0, z)$. Let us first look at the half plane $x \leq 0, -\infty < z < \infty$. We have here

$$E_v(x, z) = -\int_0^\infty E_v(0, z') \frac{\partial G^{(1)}}{\partial x'}(x, z, 0, z')\, dz', \tag{2.1}$$

where

$$G^{(1)}(x, z, x', z') = \frac{i}{4}\left[H_0^{(1)}\{k[(x - x')^2 + (z - z')^2]^{1/2}\}\right.$$

$$\left. - H_0^{(1)}\{k[(x + x')^2 + (z - z')^2]^{1/2}\}\right]$$

and $H_0^{(1)}$ is the usual notation for the Hankel function. This Green's function is quite similar to the one which we employed in Part I. However, we note that we have subtracted a source free term from the first term and this has the effect of making $G^{(1)}$ vanish for $x = 0$. Thus, noting boundary conditions for $E_v(x, z)$ and $G^{(1)}$ on the line $x = 0$, we are left with (2.1).

As for the strip, we choose a Green's function $G^{(2)}(x, z, x', z')$ which vanishes for $x = 0$ and whose normal derivative vanishes for $x = a/2$. Such a Green's function has the form

$$G^{(2)}(x, z, x', z') = \frac{2}{\pi}\sum_{n=1}^{\infty} \sin\frac{(2n + 1)\pi x}{a} \sin\frac{(2n + 1)\pi x'}{a} \exp\{-|z - z'|\pi\kappa_n/a\}/\kappa_n$$

$$+ \frac{2i}{\kappa a} \sin\frac{\pi x}{a} \sin\frac{\pi x'}{a} \exp\{i\kappa|z - z'|\}$$

where

$$\kappa_n = \{(2n + 1)^2 - (ak)^2/\pi^2\}^{1/2}.$$

Upon applying Green's theorem, with $G^{(2)}$ as a kernel, to the strip $-\infty < z < \infty$, $0 \leq x \leq a/2$ and noting the form of $E_v(x, z)$ for $|z| \to \infty$, we get

$$E_v(x, z) = \int_0^\infty E_v(0, z') \frac{\partial G^{(2)}}{\partial x'}\, dz' + \rho_1 \sin\frac{\pi x}{a} e^{i\kappa z}, \tag{2.2}$$

where $G^{(2)}$ has been evaluated at $x' = 0$.

Equation (2.2) contains the definition of the reflection coefficient which we desire. For z large and negative $E_v(x, z)$ is asymptotic to $(\rho_1 e^{i\kappa z} + \rho_2 e^{-i\kappa z}) \sin(\pi x/a)$. But from Eq. (2.2), we see that this is equal to

$$\rho_1 \sin\frac{\pi x}{a} e^{i\kappa z} + \frac{2i\pi}{a^2\kappa}\int_0^\infty E_v(0, z') \sin\frac{\pi x}{a} e^{i\kappa(z' - z)}\, dz'.$$

From this it follows that

$$\rho_2 = \frac{2i\pi}{a^2\kappa}\int_0^\infty E_y(0, z')e^{i\kappa z'}\,dz'.$$

We note that ρ_2 is the unilateral Fourier transform of $E_y(0, z)$ evaluated at $-\kappa$. In view of the fact that we shall solve our problem by Fourier methods, this quantity will appear directly.

We can now form the desired integral equation by observing that the z component of the magnetic field is continuous on the surface $x = 0$, $z > 0$. On this surface we have then

$$\int_0^\infty E_y(0, z')\left[\frac{\partial^2 G^{(1)}}{\partial x\,\partial x'} + \frac{\partial^2 G^{(2)}}{\partial x\,\partial x'}\right]dz' + \frac{\rho_1\pi}{a}e^{i\kappa z} = 0, \tag{2.3}$$

where x and x' are now evaluated at zero. Eq. (2.3) is of the Wiener-Hopf type because of the limits on the integral and the particular z dependence of the Green's function. For analytical convenience, we assume that k has a small positive imaginary part.

3. The Fourier transform solution of Eq. (2.3). Following Wiener and Hopf we extend Eq. (2.3) for all z to read

$$\phi(z) = \int_{-\infty}^\infty E_y(0, z')\left[\frac{\partial^2 G^{(1)}}{\partial x\,\partial x'} + \frac{\partial^2 G^{(2)}}{\partial x\,\partial x'}\right]dz' + \phi_0(z), \tag{3.1}$$

where the x and x' in the Green's functions are evaluated at zero. We define

$$E_y(0, z) \equiv 0 \qquad z < 0,$$

$$\phi_0(z) \equiv \frac{\pi\rho_1}{a}e^{i\kappa z} \qquad z > 0,$$

$$\equiv 0 \qquad z < 0,$$

and

$$\phi(z) \equiv 0 \qquad z > 0.$$

It is a simple matter to find the growth of $E_y(0, z)$ for $z \gg 0$, and $\phi(z)$ for $z \ll 0$. We shall verify with the solution of the problem that they are both integrable for finite z. $E_y(0, z)$ is asymptotic to $e^{ikz}/z^{1/2}$ for $z \gg 0$, while $\phi(z)$ is asymptotic to either $e^{-ikz}/z^{1/2}$ or $e^{-i\kappa z}$ depending upon which goes to zero more slowly for $z \ll 0$.[2] Furthermore $G^{(1)}(0, z, 0, z')$ is asymptotic to e^{ikz} for $z \gg z'$ and e^{-ikz} for $z \ll z'$, at least insofar as the exponential growth is concerned, while $G^{(2)}(0, z, 0, z')$ is asymptotic to $e^{i\kappa z}$ or $e^{-i\kappa z}$ depending on whether $z \gg z'$ or $z \ll z'$. With this information at hand, we can define the regions of regularity of the various Fourier transforms we shall encounter.

Consider first, the unilateral Fourier transform of $E_y(0, z)$,

$$\psi_1(w) = \int_0^\infty e^{-iwz}E_y(0, z)\,dz.$$

$\psi_1(w)$ is regular in the lower half plane $\Im m w < \Im m k$ because of the growth of $E_y(0, z)$ for $z \gg 0$. As for $\phi(z)$, we note that

[2] With appropriate restrictions on the imaginary part of k, we may determine the magnitude of $\Im m k$ relative to $\Im m\,\kappa$. However, as we shall see, the information is unessential.

$$\psi_2(w) = \int_{-\infty}^{0} e^{-iwz}\phi(z)\,dz$$

is regular in the upper half plane $\Im m w > -\Im m k$ (or $-\Im m \kappa$). The right side of this inequality is the smaller of the two quantities $(\Im m k, \Im m \kappa)$. The expression $G^{(1)}(0, z, 0, z')$ will have the bilateral Fourier transform

$$g_1(w) = \int_{-\infty}^{\infty} e^{-iwz}G^{(1)}(0, z, 0, z')\,dz$$

which is regular in the strip $-\Im m k < \Im m w < \Im m k$, while the bilateral transform of $G^{(2)}(0, z, 0, z')$,

$$g_2(w) = \int_{-\infty}^{\infty} e^{-iwz}G^{(2)}(0, z, 0, z')\,dz$$

is regular in the strip $-\Im m \kappa < \Im m w < \Im m \kappa$. Upon noting the transform of $\phi_0(z)$, we see that all transforms involved in Eq. (3.1) have a common strip of regularity, $-\Im m k$ (or $-\Im m \kappa$) $< \Im m w < \Im m k$ (or $\Im m \kappa$) and it is thus permissible to apply the Fourier transform to Eq. (3.1).

We have given elsewhere the Fourier transforms of the Green's functions. For example, with $\omega = (k^2 - w^2)^{1/2}$, we have

$$\int_{-\infty}^{\infty} e^{-iwz}G^{(1)}(x, z, x', z')\,dz = -\frac{1}{\omega}\exp\{-iwz' - i\omega x'\}\sin \omega x, \qquad \text{for } x > x',$$

$$= -\frac{1}{\omega}\exp\{-iwz' - i\omega x'\}\sin \omega x', \quad \text{for } x < x',$$

and this is clearly regular in the strip $-\Im m k < \Im m w < \Im m k$. For the transform of $G^{(2)}(x, z, x', z')$, we have

$$\int_{-\infty}^{\infty} e^{-iwz}G^{(2)}(x, z, x', z')\,dz = \frac{e^{-iwz'}\sin \omega x' \cos \omega(x - a/2)}{\omega \cos (a\omega/2)} \qquad \text{for } x > x',$$

$$= \frac{e^{-iwz'}\sin \omega x \cos \omega(x' - a/2)}{\omega \cos (a\omega/2)} \qquad \text{for } x < x',$$

and this is regular in the strip $-\Im m \kappa < \Im m w < \Im m \kappa$.

If we now apply the Fourier transform theorem to Eq. (3.1) we get immediately

$$\psi_2(w) = \psi_1(w)\omega[\tan (a\omega/2) + i] + \pi\rho_1/ai(w - \kappa)$$

$$= i\omega\psi_1(w)\frac{\exp(-ia\omega/2)}{\cos a\omega/2} + \pi\rho_1/ai(w - \kappa).$$

Eq. (3.2) may now be split into two parts, one of which is regular in the appropriate lower half plane, while the other of which is regular in the appropriate upper half plane. We have already described these details in Part I, so without any further discussion we can write

$$\pi\rho_1 K_+(\kappa)/ai(w - \kappa) + i\psi_1(w)K_-(w) = E(w) \tag{3.3}$$

and

$$\pi\rho_1[K_+(w) - K_+(\kappa)]/ai(w - \kappa) - \psi_2(w)K_+(w) = E(w) \tag{3.4}$$

$E(w)$ is an integral function and

$$K_-(w) = N_-(w)/D_-(w),$$

where

$$N_-(w) = \{k - w\}^{1/2} \exp\left[-\frac{a\omega i}{\pi} \arctan\left\{\frac{k+w}{k-w}\right\}^{1/2} + \chi(w)\right]$$

and

$$D_-(w) = (w - \kappa) \prod_{n=1}^{\infty} [\{1 - (ak)^2/\pi^2(2n+1)^2\}^{1/2} + iaw/\pi(2n+1)] \exp\{-iaw/\pi(2n+1)\}$$

while

$$\frac{1}{K_+(w)} = -\frac{\pi^2}{a^2} K_-(-w).$$

$\chi(w)$ is an integral function which is chosen so as to render $K_-(w)$ and $K_+(w)$ of algebraic growth in the appropriate half plane as $|w|$ becomes infinite. We have already seen that

$$\chi(w) = \frac{iaw}{2\pi}\left[\log\frac{i\pi}{ak} + 3 - \gamma\right]$$

where γ is the Euler-Mascheroni constant. With $\chi(w)$ so chosen, $\Im m w$ in the appropriate lower half plane and $|w| \gg 0$, $K_-(w) = O(w^{1/2})$. Upon employing these asymptotic forms and noting that $\psi_1(w)$ and $\psi_2(w)$ approach zero for $|w| \gg 0$ with w in the appropriate half plane, we find by a direct application of Liouville's theorem, that $E(w)$ is identically zero. Thus we have

$$\psi_1(w) = \int_0^{\infty} e^{-iwz} E_y(0, z)\, dz = \frac{\pi\rho_1 K_+(\kappa)}{a(w-\kappa)K_-(w)}.$$

This tells us, incidentally, that $\psi_1(w) = O(w^{-3/2})$ for $|w| \gg 0$, $\Im m w$ in the appropriate lower half plane, and hence that $E_y(0, z) = O(z^{1/2})$, $z \to 0^+$.

It is now a simple task to compute the reflection coefficient. We have that

$$\rho_2 = \frac{2i\pi}{a^2\kappa}\psi_1(-\kappa) = \frac{2i\pi}{a^2\kappa}\int_0^{\infty} e^{i\kappa z} E_y(0, z)\, dz.$$

Hence $\rho_2/\rho_1 = R$, the reflection coefficient, is

$$R = -\frac{i\pi^2}{a^3\kappa^2}\frac{K_+(\kappa)}{K_-(-\kappa)}.$$

Let us now dispense with the assumption that k has an imaginary component. We find that

$$R = i\left\{\frac{k-\kappa}{k+\kappa}\right\}^{1/2} \exp\left[2i\Theta_1 + 2i \arctan\left\{\frac{k-\kappa}{k+\kappa}\right\}^{1/2} + 2\chi(\kappa)\right]$$

where Θ_1 has been defined in Part I. Here $\pi < ak < 2\pi$, and $a^2\kappa^2 = a^2k^2 - \pi^2$.

4. Some remarks on the method employed. We have discussed and solved in this series of papers, a group of free space problems in electromagnetic theory. Because of the peculiar geometry, it was always possible to formulate this set of problems as integral equations which are closely related to the Wiener-Hopf type. Despite the subtle difference which exists between the integral equation we have treated and the original Wiener-Hopf theory, the mathematical machinery carries over. It is worth noting here that the class of integral equations we solve here, for example Eq. (3.1), belongs to the inversion formula type. We shall pursue this remark elsewhere.

It is clear that in order to apply Fourier methods to this class of integral equations we must be certain that the various functions involved have proper growths at infinity, as well as in any finite interval. An examination of the integral equations we have solved reveals that integrability over any finite interval is demanded if the equations have been properly formulated. The same remark holds true for conditions at infinity. Indeed, on the basis of the integrals we use to calculate certain physical parameters, the integrability condition at the origin would enter immediately. As for the asymptotic form of the various field quantities, the physics of the situation dictates the precise form that we require. Every Fourier transform solution we have obtained can be readily shown to have appropriate properties and therefore the field quantities also do.

DIABATIC FLOW OF A COMPRESSIBLE FLUID*

BY

BRUCE L. HICKS

Ballistic Research Laboratories, Aberdeen Proving Ground

I. Introduction

The principal concern of theoretical aerodynamics in the past has been with the theory of adiabatic flow. The development of a corresponding theory of diabatic[1] flow is needed in order to provide proper theoretical interpretations of present basic experimental studies of combustion aerodynamics and to make possible the most general application of these studies to the design of combustion and ignition apparatus.

Quite general discussions have been given of several features of the diabatic flow of a compressible viscous fluid. Thus Kiebel [K-1]** has classified such flow into a number of dynamically permissible types with applications to meteorology, and Bateman [B-1] and others have proposed several variational principles. Unsteady diabatic flows less general than those studied by Kiebel are considered in meteorology but the heat addition function apparently does not usually enter explicitly. (See for example the development of the Bjerknes theorem in Ch. VI of Haurwitz [H-7].) In the one-dimensional or hydraulic approximation, steady, frictional, continuous (i.e., shockless) diabatic flow ([H-5] and references cited there) and discontinuous steady flow (deflagrations and detonations—see [B-1], [B-2], [H-1], and references in [B-2], [H-1], and [H-5]) have both been treated. However, there appears to be no published analysis which is designed to apply specifically to steady, diabatic flow of an inviscid, compressible fluid in two and three dimensions, although some transformations of the equations are useful in both adiabatic and diabatic flow.

Our purpose in the present series of papers is to describe the principal characteristics of steady diabatic compressible flow. The treatment will apply to both thin and thick burning regions and to subsonic and supersonic flow conditions if it is understood that algebraic equations should be added whenever necessary to express conservation of energy, mass, and momentum across flow discontinuities. The theory will not apply generally where viscous effects may be important, as in the interior of detonation fronts, ignition within boundary layers, etc. Our discussion depends upon the use of vectors other than the velocity vector to represent the flow, these vectors being chosen in order that the complexity of the basic differential equations can be reduced. A general discussion of these vectors will be given later (reference H-6). In the present paper, we shall derive new basic equations in terms of the velocity vector and in two other vector languages. We shall exhibit the physical content and implications of these equations for rotational and irrotational flow, sub- and supersonic flow and compare them with their counterparts in adiabatic flow.[2]

No effort is made to solve specific technical problems in this paper because it is believed desirable first to survey the whole field in coherent fashion. Formal manipula-

---

*Received April 29, 1947.

[1]The word diabatic in place of nonadiabatic was suggested by Dr. D. J. Montgomery.

**The number in square brackets refers to the list of references at the end of the paper.

[2]Some of these results were presented first in [H-2], [H-3].

tion of the equations is based upon simple physical ideas. Boundary and initial-value problems are not discussed. Although only a few types of quasi-linear partial differential equations occur in this diabatic theory, and all these types have appeared in adiabatic theory, the boundary-value problems may be more difficult than in adiabatic flow because of the interdependence of the flow pattern and the position of an extended burning region.

Projected applications of the diabatic theory in the field of combustion aerodynamics provided the initial impetus for our research. It is hoped that the theory may prove useful in other fields as well.

II. Three-Dimensional Flow

1. Fundamental equations. Formulation of the diabatic flow equations will be based upon the ordinary hydrodynamic equations of continuity and motion for an inviscid fluid

$$\nabla \cdot \rho \mathbf{V} = 0, \tag{1.1}$$

$$\nabla p + \rho \mathbf{V} \cdot \nabla \mathbf{V} = 0, \tag{1.2}$$

(where p, ρ and $\mathbf{V}$ represent static pressure, density, and the velocity vector), the first law in the form given by Vazsonyi [V-2]

$$Q = c_p \mathbf{V} \cdot \nabla T_t , \tag{1.3}$$

(where Q is the heat added to the fluid per unit mass and time and $c_p T_t$ is the stagnation or total enthalpy), and the equation of state for a perfect gas

$$p = R\rho T, \tag{1.4}$$

(where T is the static temperature and R, the gas constant).

These equations can be modified to advantage if a vector $\mathbf{W}$ is introduced which was found to be appropriate by Crocco [C-1] (compare [T-1]) for adiabatic, iso-energetic flow and in reference [H-4] for the most general type of adiabatic compressible flow. This vector $\mathbf{W}$ is defined as

$$\mathbf{W} = \mathbf{V}/V_t , \tag{1.5}$$

where

$$V_t = (2c_p T_t)^{1/2} \tag{1.6}$$

is the "limiting velocity" at any point in the flow. The quantity W^2 thus represents the ratio of the kinetic energy of a fluid particle to its total energy (cf. ref. [H-4] and appendix of [H-5]). It is noted that W and M, the local Mach number, are related by the equation

$$(1 - W^2)\left(1 + \frac{\gamma - 1}{2} M^2\right) = 1. \tag{1.7}$$

A second vector transformation that will be used is the substitution $\mathbf{V} = a\mathbf{M}$, where

$$a = (\gamma R T)^{1/2} \tag{1.8}$$

is the local velocity of sound.

In terms of W, the total or stagnation temperature T_t and total pressure p_t are given by

$$T_t = T(1 - W^2)^{-1} = T\left(1 + \frac{\gamma - 1}{2} M^2\right), \tag{1.9}$$

$$p_t = p(1 - W^2)^{-\gamma/\gamma-1} = p\left(1 + \frac{\gamma - 1}{2} M^2\right)^{\gamma/\gamma-1}. \tag{1.10}$$

The symbols ω_V , ω_W , and ω_M will be used for the vorticity functions $\nabla \times \mathbf{V}$, $\nabla \times \mathbf{W}$ and $\nabla \times \mathbf{M}$.

2. Transformed equations. The basic dynamical equations, (1.1) to (1.4), can be reduced to a smaller number of equations containing only $\mathbf{W}$, ω_W , p_t , and a parameter q proportional to Q. We first express the continuity Eq. (1.1) in terms of $\mathbf{W}$, p_t , and T_t with the help of equations (1.4), (1.9), (1.10).

$$\nabla \cdot \mathbf{W} + \mathbf{W} \cdot [\nabla \log (p_t T_t^{-1/2}) + (\gamma - 1)^{-1} \nabla \log (1 - W^2)] = 0. \tag{2.1}$$

Similarly, using Eq. (1.10), the definition of ω_W and the identity $\mathbf{W} \cdot \nabla \mathbf{W} = \frac{1}{2} \nabla W^2 - \mathbf{W} \times \omega_W$, we find that the equation of motion (1.2) becomes

$$\nabla \log p_t = \frac{2\gamma}{\gamma - 1} (1 - W^2)^{-1} [\mathbf{W} \times \omega_W - \tfrac{1}{2} \mathbf{W} (\mathbf{W} \cdot \nabla \log T_t)]. \tag{2.2}$$

We now introduce the factor q which is defined as

$$q = Q/V_t^3 \tag{2.3}$$

and which therefore satisfies, according to Eqs. (1.3), (1.5), the relation

$$q = \tfrac{1}{2} \mathbf{W} \cdot \nabla \log T_t = \mathbf{W} \cdot \nabla \log V_t . \tag{2.4}$$

(The quantity $q_W = q/(1 - W^2)$ will also be used later; q_W corresponds to the notation of [H-6].)

Equation (2.2) now becomes

$$\nabla \log p_t = \frac{2\gamma}{\gamma - 1} (1 - W^2)^{-1} [\mathbf{W} \times \omega_W - q\mathbf{W}] \tag{2.5}$$

which we will call the $\mathbf{W}$ equation of motion. This equation is a generalization of the Bjerknes type equation (cf. [V-2], Eq. (6.1))

$$-T \nabla S + c_p \nabla T_t = \mathbf{V} \times \omega \tag{2.6}$$

in $\mathbf{V}$ language and of the $\mathbf{W}$ equation previously given in [H-4].

Final reduction of the continuity equation is effected by calculating $\mathbf{W} \cdot \nabla \log (p_t T_t^{-1/2})$ from Eqs. (2.4), (2.5) and substituting in Eq. (2.1):

$$\nabla \cdot \mathbf{W} + \frac{1}{\gamma - 1} \mathbf{W} \cdot \nabla \log (1 - W^2) = q\left(1 + \frac{\gamma + 1}{\gamma - 1} W^2\right)/(1 - W^2) \tag{2.7}$$

or

$$\nabla \cdot (1 - W^2)^{1/\gamma-1} \mathbf{W} = q\left(1 + \frac{\gamma + 1}{\gamma - 1} W^2\right)(1 - W^2)^{(2-\gamma)/(\gamma-1)}. \tag{2.8}$$

When $q = 0$, the equation

$$\nabla\cdot(1 - W^2)^{1/\gamma-1}\mathbf{W} = 0 \tag{2.9}$$

is obtained (cf. [C-1] and [H-4]).

Similar development in the **M** language leads to two equations analogous to (2.5) and (2.8) (cf. [H-4]):

$$\frac{1}{\gamma}\nabla \log (pe^{\gamma M^2/2}) - \frac{\gamma - 1}{\gamma + 1}\mathbf{M}\nabla\cdot\mathbf{M} = \frac{2}{\gamma + 1} q_M\mathbf{M} + \mathbf{M} \times \omega_M , \tag{2.10}$$

$$\nabla\cdot\left(1 + \frac{\gamma - 1}{2} M^2\right)^{-(\gamma+1)/2(\gamma-1)}\mathbf{M} = (1 + \gamma M^2)\left(1 + \frac{\gamma - 1}{2} M^2\right)^{-(3\gamma-1)/2(\gamma-1)} q_M , \tag{2.11}$$

where $q_M = (\gamma - 1)Q/2a^3$ in accordance with the convention to be used in [H-6]. It is important to note that the dimensionless quantities q_W and q_M rather than Q appear in these transformed equations. As will be shown in H-6, reductions of the form employed here, (i.e., $\mathbf{V} = [g(N)RT]^{1/2}\,\mathbf{N}$), never yield equations in which Q is the most appropriate heating factor.

3. Comparison with adiabatic flow. In adiabatic flow, T_t and p_t are constant on streamlines though they can vary between streamlines [H-4]. In diabatic flow, however, the variation of T_t on streamlines is prescribed by Eq. (1.3) or (2.4), its variation normal to streamlines still not being restricted. The rate of change of p_t along a streamline is proportional to qW and normal to streamlines, to the appropriate component of $\mathbf{W} \times \omega_W$ as in adiabatic flow [H-4]. The **W** language is thus particularly useful because it permits simple expression of the variations in stagnation pressure. In the M language, the quantity $pe^{\gamma M^2/2}$ is a more appropriate variable according to Eq. (2.10) than p or p_t , but its variation along streamlines depends both upon q_M and upon $\nabla\cdot M$.

We note that the new equations of motion (2.5), of continuity (2.7), and the first law (2.4) are more compact than the conventional equations (1.1), (1.2), (1.3) having been stripped of a non-essential variable in that now only p_t , q, and W appear in place of p, ρ, Q and **V**. The new equations also connect directly the changes in p_t and q with variation throughout the **W** field. By taking the curl of Eq. (2.5) we can eliminate p_t and find another differential equation in addition to Eq. (2.7) which describes the variation of the **W** field. Thus,

$$\nabla \times [\mathbf{W} \times \omega_W/(1 - W^2)] = \nabla \times [q\mathbf{W}/(1 - W^2)] \tag{3.1}$$

which is a generalization of Crocco's equation [C-1] and also of an equation in reference [H-4]. (Cf. later discussion of $\omega_V = 0$ and of $\omega_W = 0$.) It is noted that q in Eq. (3.1) can be eliminated by use of Eq. (2.8) resulting in a differential equation involving **W** only. The analog of (3.1) in **M** language will contain terms in $\nabla \times (\mathbf{M}\ \nabla\cdot\mathbf{M})$ which have no counterpart in the **W** language. Elimination of q_M with the help of (2.11) is possible however.

The behavior of the fluid when it reaches sonic velocity in a diabatic flow is also worth examination. Let us use the convention suggested in [H-4] that $\nabla\cdot(\mathbf{W}/W) = \nabla\cdot\mathbf{s}$ is a measure of the fractional rate of variation along a streamline of stream-tube area. Then from Eq. (2.7)

$$\nabla\cdot\mathbf{s} + (1 - W^2)^{-1}\left(1 - \frac{\gamma+1}{\gamma-1}W^2\right)\mathbf{s}\cdot\nabla \log W = qW^{-1}(1 - W^2)^{-1}\left(1 + \frac{\gamma+1}{\gamma-1}W^2\right). \tag{3.2}$$

For local velocity equal to the local velocity of sound, $W = [(\gamma - 1)/(\gamma + 1)]^{1/2}$ and if the flow is continuous,

$$\nabla\cdot\mathbf{s} = q(\gamma + 1)^{3/2}(\gamma - 1)^{-1/2} \tag{3.3}$$

Thus the stream-tube diverges at sonic velocity if $q > 0$, converges if $q < 0$ and has minimum area for the adiabatic case $q = 0$ (cf. discussion in H-4). "One-dimensional" analogs of Eq. (3.3) are described in H-5.

4. Types of Irrotational flow: $\omega_V = 0$. We first observe the consequences of irrotationality in the **V** field. Just as in adiabatic flow we find at once from Eq. (1.2) (cf. [H-4]) that for $\omega_V = 0$, the flow must be barytropic, i.e., that $p = F(\rho)$. The continuity equation (1.1) then leads us to a partial differential equation for a velocity potential φ_V of familiar form. Thus from Eq. (1.2)

$$\mathbf{V}\cdot\nabla \log \rho = -a_b^{-2}\mathbf{V}\cdot\nabla \tfrac{1}{2}V^2, \tag{4.1}$$

where $a_b^2 = dp/d\rho$ and Eq. (1.1) can be written in summation notation as follows:

$$\sum_i (a_b^2 - V_i^2)\frac{\partial V_i}{\partial x_i} - 2\sum_{i>j} V_i V_j \frac{\partial V_i}{\partial x_j} = 0, \tag{4.2}$$

where V_r and x_r are the components of velocity and position vectors, respectively. Placing $V_i = \partial\varphi_v/\partial x_i$ we find

$$\sum_i \left[a_b^2 - \left(\frac{\partial\varphi_V}{\partial x_i}\right)^2\right]\frac{\partial^2\varphi_V}{\partial x_i^2} - 2\sum_{i>j}\frac{\partial\varphi_V}{\partial x_i}\frac{\partial\varphi_V}{\partial x_j}\frac{\partial^2\varphi_V}{\partial x_i\,\partial x_j} = 0. \tag{4.3}$$

Equation (4.3) changes type from elliptic to hyperbolic as V increases and passes through the value a_b. When $M \neq 1$, a_b can be greater or less than a, but for $M = 1$, $a_b = a$ and change of type still corresponds to passage between subsonic and supersonic flow. This correspondence no longer holds for some types of irrotational diabatic flow [cf. Eqs. (4.11), (4.13); Sec. 6; and Eq. (5.1)].

Further development yields relations between a_b and Q. Vazsonyi [V-2] gives the equation governing the entropy variation, ∇S, in the form

$$\mathbf{V}\cdot\nabla S = Q/T \tag{4.4}$$

which, together with the definition of S

$$dS = c_v d \log (\rho^{-\gamma}p), \tag{4.5}$$

leads to

$$\left(\frac{1}{\gamma}\frac{d \log p}{d \log \rho} - 1\right)\mathbf{V}\cdot\nabla \log \rho = Q/c_pT \tag{4.6}$$

and finally, with introduction of φ_v to

$$\left(\frac{1}{\gamma}\frac{d \log p}{d \log \rho} - 1\right)\nabla^2\varphi_V = -Q/c_pT. \tag{4.7}$$

We see that Q/T is a measure of the divergence of the **V** field and of the departure of the (p, ρ) relation from the adiabatic one.

This result can be thrown in a form that is directly comparable with the one-dimensional theory. From Eqs. (1.1), (1.2) for $\omega_V = 0$

$$\nabla \cdot \mathbf{s} = -\frac{\partial \log \rho}{\partial s} - \frac{\partial \log V}{\partial s}, \tag{4.8}$$

$$\frac{\partial p}{\partial s} = -\rho V^2 \frac{\partial \log V}{\partial s}, \tag{4.9}$$

where $\partial/\partial s = \mathbf{s} \cdot \nabla$. Hence

$$\nabla \cdot \mathbf{s} = \left(\frac{1}{\gamma M^2} \frac{d \log p}{d \log \rho} - 1\right) \frac{\partial \log \rho}{\partial s}. \tag{4.10}$$

We can then use Eq. (4.6) to eliminate $\partial \log \rho / \partial s$ obtaining

$$\frac{dp}{d\rho} = a_b^2 = a^2 M^2 (\nabla \cdot \mathbf{s} - Q/c_v TV)/(M^2 \nabla \cdot \mathbf{s} - Q/c_v TV). \tag{4.11}$$

When $Q = 0$, we obtain the usual adiabatic formula $dp/d\rho = a^2$. We now see that, depending upon the relative magnitudes of M, $\nabla \cdot \mathbf{s}$ and $Q/c_v TV$, a_b^2 may differ from a^2 by a factor of either sign as well as any magnitude (cf. S-1). In comparing with one-dimensional flow in the direction of x in a duct of (variable) area, we place $\nabla \cdot \mathbf{s} = -d\alpha/dx$ and $Q/(c_v TV) = d\theta^*/dx$ (see reference [H-5]) and obtain

$$\frac{dp}{d\rho} = a^2 M^2 (d\alpha + d\theta^*)(M^2\, d\alpha + d\theta^*)^{-1} \tag{4.12}$$

which is what could also be computed from Eqs. (15) and (17) of this reference.

It is noted that the analog of (3.2) in **V** language is

$$\nabla \cdot \mathbf{s} = (Q/c_v TV) - (1 - M^2) \frac{\partial \log V}{\partial s} \tag{4.13}$$

which reduces to the known expression for adiabatic flow $\nabla \cdot \mathbf{s} = -(1 - M^2)(\partial \log V)/\partial s$ when $Q = 0$ and to expression (3.3) for $M = 1$.

We now show that given $F(\rho)$ and appropriate boundary conditions Q can be determined as a function of position. Integration of Bernoulli's equation gives the functional relationship between V and ρ.

$$\frac{1}{2} V^2 = \frac{1}{2} V_0^2 - \int \rho^{-1} F'(\rho)\, d\rho \tag{4.14}$$

from which $a_b^2 = F'(\rho)$ is also expressible implicitly as a function of $V = |\nabla \varphi_V|$. Accordingly, the continuity Eq. (4.3) is a quasi-linear partial differential equation which can be solved with appropriate boundary conditions for φ_V as a function of position. Subsequent calculation of $\nabla^2 \varphi_V$ and $(d \log p)/(d \log \rho)$ yields Q/T as a function of V along any streamline. The energy equation (1.3) can be written for any one streamline s_1 as follows:

$$\frac{dT}{ds_1} - (Q/c_v TV)T = -\frac{1}{2c_v} \frac{dV^2}{ds_1}, \tag{4.15}$$

where V^2 is a known function of s_1. Accordingly, T as a function of s_1, and therefore also Q can be found along each streamline and therefore throughout the field of flow. This process, although complex, is direct; it would be essentially more difficult to proceed in the opposite direction, that is, with knowledge of Q (or some combination involving Q) because compatible solutions of (4.3) and (4.7) would not be obtainable for arbitrary Q functions.

A note on the place of the Earnshaw pressure-density relation in the theory is in order. When $\omega_v = 0$ in diabatic flow, a hodograph transformation yields an additional term proportional to Q in the hodograph partial differential equation for ψ, the stream function. Assumption of a linearized (i.e., Earnshaw) p vs. ρ^{-1} relation corresponds to a restriction upon the type of Q function that can be specified. In order to obtain the partial differential equation for ψ used by Chaplygin and by von Kármán, the additional term proportional to Q must be neglected. Therefore, treatments which utilize *any* linearized p vs. ρ^{-1} relation in the Chaplygin equation can be regarded as approximate treatments of a special type of diabatic flow. (Compare also [V-1], p. 348.)

The previous discussion also indicates that a Glauert-Prandtl treatment (cf. [T-2]) should be feasible. In the series of approximations leading to the expansion or contraction factor $(1 - M_b^2)^{1/2}$ where $M_b = V_0/a_{b0}$, a_b will now replace a. The method is especially simple when the p, ρ relation is polytropic, $p \propto \rho^k$. Then

$$\left(1 - \frac{k}{\gamma}\right)\nabla^2\varphi_V = Q/c_pT \tag{4.16}$$

and for small distortions of the flow

$$\nabla^2\varphi_V = M_b^2\frac{\partial^2\varphi_V}{\partial x^2}. \tag{4.17}$$

Combination of these equations yields

$$\frac{\partial^2\varphi_V}{\partial x^2} = M_b^{-2}\left(1 - \frac{k}{\gamma}\right)^{-1}Q/c_pT. \tag{4.18}$$

Now transform the variables according to the scheme

$$x = (1 - M_b^2)^{1/2}\xi, \qquad y = \eta, \qquad z = \zeta, \qquad \varphi_V(x, y, z) = \varphi_V^*(\xi, \eta, \zeta)$$

The equations (4.17), (4.18) become

$$\nabla^2_{\xi,\eta,\zeta}\varphi_V^* = 0 \tag{4.19}$$

$$\frac{\partial^2\varphi_V^*}{\partial\xi^2} = (1 - M_b^2)Q/c_pTM_b^2\left(1 - \frac{k}{\gamma}\right) \tag{4.20}$$

According to the second of these equations,

$$\nabla^2_{\xi,\eta,\zeta}\frac{Q}{c_pT} = 0 \tag{4.21}$$

Thus in the Glauert-Prandtl approximation in a diabatic, polytropic field of flow, the function Q/c_pT must be harmonic in the variables (ξ, η, ζ). As $M_b \to 0$, Q and $\nabla^2\varphi_V$ likewise approach zero.

Other assumed (p, ρ) relations than polytropic or isentropic have not so far led to simple results of interest.

5. Types of irrotational flow (cont'd.): $\omega_W = 0$. In an irrotational **W** field, $\omega_W = 0$, and the equation of motion (2.5) becomes

$$\nabla \log p_t = -\frac{2\gamma}{\gamma - 1} q\mathbf{W}/(1 - W^2). \tag{5.1}$$

Thus the only change in stagnation pressure is along streamlines and is of "momentum pressure drop" type due to the heating. If this equation is integrable

$$\log p_t = -\frac{2\gamma}{\gamma - 1} \int q_W(\varphi_W)\, d\varphi_W ,$$
$$q = (1 - W^2) q_W(\varphi_W), \tag{5.2}$$

in accordance with the notation to be used in [H-6], which restricts the possible modes of variation of q or q_W in that now q_W only varies along streamlines and must therefore be a function of φ_W . Accordingly, the W continuity equation can be written as follows:

$$\nabla \cdot (1 - W^2)^{1/\gamma - 1}\mathbf{W} = \left(1 + \frac{\gamma + 1}{\gamma - 1} W^2\right)(1 - W^2)^{1/\gamma - 1} q_W(\varphi_W). \tag{5.3}$$

In terms of φ_W this becomes

$$\sum_i \left[1 - W^2 - \frac{2}{\gamma - 1}\left(\frac{\partial \varphi_W}{\partial x_i}\right)^2\right] \frac{\partial^2 \varphi_W}{\partial x_i^2} - \frac{4}{\gamma - 1} \sum_{i>j} \frac{\partial \varphi_W}{\partial x_i} \frac{\partial \varphi_W}{\partial x_j} \frac{\partial^2 \varphi_W}{\partial x_i\, \partial x_j}$$
$$= \left(1 + \frac{\gamma + 1}{\gamma - 1} W^2\right)(1 - W^2) q_W(\varphi_W). \tag{5.4}$$

This quasi-linear partial differential equation changes type when $W^2 = \sum_i (\partial \varphi_W / \partial x_i)^2 = (\gamma - 1)/(\gamma + 1)$ (i.e., for $M = 1$) from elliptic to hyperbolic as W increases. Equation (5.4) is reminiscent of Crocco's equation (number 11 in [C.1]) in the stream function for two-dimensional rotational flow in that a function (arbitrary) of the dependent variable appears on the R H S. Other similar equations will be derived later in generalizing the Crocco theory. It appears that equation (5.4) might be easier to solve than its analog for the irrotational **V** field, Eq. (4.3). The key functions in the two cases are $q_W(\varphi_W)$ and $a_b(V^2)$ where $V^2 = \sum_i (\partial \varphi_V / \partial x_i)^2$ and the second function involves higher order derivatives of the dependent variable than does the first. In both cases solution of the partial differential equation for φ is the central problem, for the equations of motion are integrated by Eqs. (4.14) and (5.2).

For the incompressible case ($W \ll 1$) Eq. (5.4) reduces to

$$\nabla^2 \varphi_W = q_W(\varphi_W) \tag{5.5}$$

which is of the same form as an equation for the potential function φ_M of the Mach vector for adiabatic flow (cf. Sec. 6; also H-4). A Glauert-Prandtl treatment of (5.4) leads to an equation similar to (5.5):

$$\nabla^2_{\xi,\eta,\zeta} \varphi^*_W = \left(1 + \frac{\gamma + 1}{\gamma - 1} W_0^2\right) q^*_W(\varphi^*_W) \tag{5.6}$$

in which now

$$x = (1 - W_0^2)^{-1/2}\left(1 - \frac{\gamma + 1}{\gamma - 1} W_0^2\right)^{1/2} \xi, \qquad y = \eta, \qquad z = \zeta,$$

$$\varphi_W(x, y, z) = \varphi_W^*(\xi, \eta, \zeta).$$

6. Types of irrotational flow (cont'd.): $\omega_M = 0$. If the **M** field is irrotational, (2.10) shows that

$$\log (pe^{\gamma M^2/2}) + \int F_M(\varphi_M)\, d\varphi_M = 0 \tag{6.1}$$

and

$$\nabla^2 \varphi_M = \frac{2\gamma}{\gamma - 1} q_M - \frac{\gamma + 1}{\gamma(\gamma - 1)} F_M(\varphi_M) \tag{6.2}$$

where $F_M(\varphi_M)$ is arbitrary.
The function $pe^{\gamma M^2/2}$ is thus constant on potential surfaces as in adiabatic irrotational flow in the **M** field [H-4].

Elimination of q_M from (2.11) and (6.2) yields the partial differential equation for φ_M :

$$\sum_i \left[1 - \gamma\left(\frac{\partial \varphi_M}{\partial x_i}\right)^2\right] \frac{\partial^2 \varphi_M}{\partial x_i^2} - 2\gamma \sum_{i>j} \frac{\partial \varphi_M}{\partial x_i} \frac{\partial \varphi_M}{\partial x_j} \frac{\partial^2 \varphi_M}{\partial x_i\, \partial x_j} = \frac{1}{\gamma}(1 + \gamma M^2) F_M(\varphi_M). \tag{6.3}$$

This equation is of the same form as (5.4), but changes type for $M = \gamma^{-1/2}$ rather than for $M = 1$. This behavior is reminiscent of a property of diabatic flow in one dimension that was described in [H-1] and [H-5]. It was there found that a compressible fluid which is subjected to heating and is flowing in a duct of constant area attained its maximum static temperature T for $M = (\gamma)^{-1/2}$. A connection between this result and the situation described by the irrotational **M** field is shown as follows. From our previous equations,

$$\mathbf{M} \cdot \nabla \log a = \frac{1}{\gamma} F_M(\varphi_M). \tag{6.4}$$

Comparison with Eq. (6.3) for the one-dimensional case $\mathbf{M} = M\mathbf{i}$ gives

$$M \frac{d \log a}{dx} = (1 - \gamma M^2)(1 + \gamma M^2)^{-1} \frac{dM}{dx} \tag{6.5}$$

and a (or T) has a maximum for $M = 1$. This maximum in the one-dimensional case occurs, of course, no matter what language is used to replace the M language. It is possible that an extremum of the temperature might also occur in three-dimensional irrotational **M** flow at $M = \gamma^{-1/2}$ if $F_M(\varphi_M)$ is chosen properly.

If throughout the flow $M \ll 1$ ("incompressible approximation") then

$$\nabla^2 \varphi_M = q_M = \frac{1}{\gamma} F_M(\varphi_M) \tag{6.6}$$

which is similar to the **W** case (5.5). If the Glauert-Prandtl type of approximation is used ($M_x = M = M_0$, $M_y \ll M_0$, $M_z \ll M_0$ where M_0 is the uniform Mach number at infinity),

$$\nabla^2_{\xi,\eta,\zeta}\varphi^*_M = \frac{1}{\gamma}(1 + \gamma M_0^2)F_M(\varphi^*_M), \tag{6.7}$$

where $x = \xi(1 - \gamma M_0^2)^{1/2}$, $y = \eta$, $z = \zeta$, $\varphi_M(x, y, z) = \varphi^*_M(\xi, \eta, \zeta)$ which is of the same form as (5.6). Note that q_M is not simply related to $F_M(\varphi_M)$ in the Glauert-Prandtl approximation for irrotational **M** flow, because $\nabla^2_{x,y,z}\varphi_M$ occurs in the second integral (6.2) of the **M** equation of motion. If $\underline{\xi}$ is put equal to $x(1 + (\gamma - 1)/2\, M_0^2)^{1/2}(1 - M_0^2)^{-1/2}$ then q_M satisfies the equation similar to

$$\nabla^2_{\underline{\xi},\eta,\zeta}\varphi^*_M = (1 + \gamma M_0^2)\left(1 + \frac{\gamma - 1}{2} M_0^2\right)^{-1} q_M \,. \tag{6.8}$$

As M_0 decreases, $\underline{\xi} \to \xi$ and (6.8) approaches Eq. (6.6) in form. For larger M_0 than corresponds to use of (6.6), $\underline{\xi}$ and ξ are still approximately equal and q_M is approximately proportional to $F(\varphi_M)$. It will be recalled that in the irrotational **W** field, $q_W = q_W(\varphi_W)$ exactly.

III. Uniplanar Flow

When we pass to the two-dimensional case of uniplanar flow we find that there are a number of systems of variables in the three languages (**V**, **W**, **M**), each of which may be useful in the proper circumstances. On the one hand, in the "physical plane" we may choose W (or M, V) and θ, the direction angle of a streamline, or alternatively, potential and stream functions φ, ψ to be the dependent variables (see [V-1]). Cartesian or curvilinear coordinates may appear as independent variables. On the other hand, in the "hodograph plane" either the Cartesian components u, v or the polar components W (or V, M), θ will enter as independent variables and usually φ, ψ as dependent variables. We will now derive or state the differential equations for uniplanar diabatic flow for a number of these cases. Their most general form, that is, when neither q nor ω vanishes, will always be given first. Because in some ways the **W** formulation appears at present to be the most convenient we will give the derivations in **W** language only.

7. The unit vectors s, n. The basic equations (2.5), (2.8) in the case of uniplanar flow are made easier to handle if the mutually orthogonal unit vectors (**s**, **n**, **k**) are introduced. The plane of the flow is normal to **k**, **s** lies in the direction of the streamline at each point and **n** in the direction normal to the streamline. The sense of the vectors is such that

$$\mathbf{k} = \mathbf{s} \times \mathbf{n}. \tag{7.1}$$

Then it follows that

$$\boldsymbol{\omega} = \omega\mathbf{k}, \tag{7.2}$$

$$\mathbf{W} \times \boldsymbol{\omega} = -W\omega\mathbf{n}. \tag{7.3}$$

Accordingly, the **W** equation of motion (2.5) becomes

$$\nabla \log p_t = -\frac{2\gamma W}{\gamma - 1}(1 - W^2)^{-1}(\omega\mathbf{M} + q\mathbf{s}). \tag{7.4}$$

With the convention

$$\frac{\partial}{\partial s} = \mathbf{s}\cdot\nabla, \qquad \frac{\partial}{\partial n} = \mathbf{n}\cdot\nabla, \tag{7.5}$$

Eq. (7.4) may be rewritten as follows:

$$\frac{\partial \log p_t}{\partial s} = -\frac{2\gamma W}{\gamma - 1}(1 - W^2)^{-1}q,$$
$$\frac{\partial \log p_t}{\partial n} = -\frac{2\gamma W}{\gamma - 1}(1 - W^2)^{-1}\omega. \tag{7.6}$$

The continuity equation is now

$$W(1 - W^2)\nabla \cdot \mathbf{s} + \left(1 - \frac{\gamma + 1}{\gamma - 1}W^2\right)\frac{\partial W}{\partial s} = \left(1 + \frac{\gamma + 1}{\gamma - 1}W^2\right)q. \tag{7.7}$$

It should be noted that $\partial/\partial s$, $\partial/\partial n$ cannot be treated like partial derivatives in general because these symbols are shorthand for the longer expressions, $\partial/h_{s'}\,\partial s'$, $\partial/h_{n'}\,\partial n'$ where $h_{s'}$ and $h_{n'}$ are elements of the metric for the orthogonal curvilinear coordinate system s', n' and $h_{s'}$, $h_{n'}$ can be functions of both s', n'. Thus p_t cannot be eliminated from (7.6) by simple cross differentiation but elimination can be effected by vector operations on Eq. (7.4) as shown in (3.1) which now can be written as

$$W(\omega\mathbf{n} + q\mathbf{s}) \times \nabla \log (1 - W^2) + \nabla \times (W\omega\mathbf{n} + q\mathbf{W}) = 0. \tag{7.8}$$

After dotting in $\mathbf{k}$ and using the unit-vector formulas in the appendix we find that

$$W\left(-\omega\frac{\partial}{\partial s} + q\frac{\partial}{\partial n}\right)\log (1 - W^2) + \omega\nabla\cdot\mathbf{W} + W\frac{\partial\omega}{\partial s} + q\omega - W\frac{\partial q}{\partial n} = 0. \tag{7.9}$$

This may be combined with (2.7) and rearranged to read

$$2W^{-1}(1 - W^2)^{-1}\left(1 + \frac{1}{\gamma - 1}W^2\right)q\omega + \omega\frac{\partial}{\partial s}\log\left[\omega/(1 - W^2)^{1/\gamma - 1}\right]$$
$$= q\frac{\partial}{\partial n}\log\left[q/(1 - W^2)\right]. \tag{7.10}$$

It is noted that this can also be written in the form

$$2(1 - W^2)^{-1}\left(1 + \frac{1}{\gamma - 1}W^2\right)q\omega + \omega\left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)\log\left[\omega/(1 - W^2)^{1/\gamma - 1}\right]$$
$$= q\left(u\frac{\partial}{\partial y} - v\frac{\partial}{\partial x}\right)\log\left[q/(1 - W^2)\right], \tag{7.11}$$

where $\mathbf{W} = u\mathbf{i} + v\mathbf{j}$.

With the help of the expression for $\mathbf{W}\cdot\nabla \log p_t$ we can find another form which can be directly compared with Crocco's results. This form is

$$2W^{-1}q\omega + \omega\frac{\partial}{\partial s}\log(\omega/p) = q\frac{\partial}{\partial n}\log\left[q/(1 - W^2)\right]. \tag{7.12}$$

Equations (7.10), (7.12) exhibit the intimate relationship between the heating factor q and the vorticity ω in steady flow. For example, if W and therefore ω are specified throughout the field of flow, then the variation of $q/(1 - W^2)$ and also of q

along directions normal to the streamlines cannot be arbitrarily specified although variation of this quantity along streamlines is still not restricted by (7.10).

Crocco's equation follows at once from (7.12) when $q = 0$, i.e.,

$$\frac{\partial}{\partial s}(\omega/p) = 0. \tag{7.13}$$

When instead, $\omega = 0$, then q is subject to the restriction (reference [H-3])

$$\frac{\partial}{\partial n}[q/(1 - W^2)] = 0. \tag{7.14}$$

This can also be written as

$$\frac{\partial q_W}{\partial n} = \frac{\partial}{\partial n}(Q/2c_pTV_t) = 0. \tag{7.15}$$

Both equations imply a degree of uniformity of the heating on curves normal to the streamlines if the flow is to be irrotational in the **W** field.

We will now suppose that W can be neglected compared to one, that is, that the flow is "incompressible." Equation (7.10) then reduces to [H-3]

$$2W^{-1}q_W\omega + \frac{\partial \omega}{\partial s} = \frac{\partial q_W}{\partial n} \tag{7.16}$$

Now when $q_W = 0$, then $\partial\omega/\partial s = 0$, a known result for rotational incompressible flow, and when $\omega = 0$, $\partial q_W/\partial n = 0$, or the factor $q = q_W$ does not vary on curves normal to streamlines. Finally, if $\partial\omega/\partial s$ and $\partial q_W/\partial n$ are to remain finite as $W \to 0$, either q_W or ω (or their product) must likewise approach zero. Very low speed continuous flow therefore cannot locally be both diabatic and rotational with arbitrarily high values of q and ω.

It is shown in the appendix that $\nabla \cdot \mathbf{s} = \partial\theta/\partial n$ where θ is the angle of inclination of a streamline. The continuity equation (7.7) accordingly becomes

$$W(1 - W^2)\frac{\partial \theta}{\partial n} + \left(1 - \frac{\gamma + 1}{\gamma - 1}W^2\right)\frac{\partial W}{\partial s} = \left(1 + \frac{\gamma + 1}{\gamma - 1}W^2\right)q \tag{7.17}$$

For $W \ll 1$,

$$\frac{\partial \theta}{\partial n} + \frac{\partial \log W}{\partial s} = q/W$$

We can complete our set of equations for uniplanar flow by adding the definition of ω, using the expression $\mathbf{k} \cdot \nabla \times \mathbf{s} = \partial\theta/\partial s$, developed in the appendix,

$$\frac{\partial \theta}{\partial s} - \frac{\partial \log W}{\partial n} = \omega/W \tag{7.19}$$

and the definition of q (energy equation)

$$\frac{\partial \log V_t}{\partial s} = q/W \tag{7.20}$$

both of which hold also when W is not $\ll 1$. These equations (7.16), (7.17), (7.19), (7.20) reduce to the conventional ones when $q = \omega = 0$.

8. Introduction of the stream and potential functions. We will follow Crocco and define a stream function in terms of $(1 - W^2)^{1/\gamma-1}\,\mathbf{W}$ rather than in terms of $\rho\mathbf{V}$ (cf., however, [T-1]), and the potential function φ will be introduced similarly. Thus, we set [H-3]

$$(1 - W^2)^{1/\gamma-1}\mathbf{W} = \nabla\varphi + \nabla \times \mathbf{k}\psi, \tag{8.1}$$

where now, for $q \neq 0$, both φ and ψ must be used. It follows that the relations

$$\frac{\partial\varphi}{\partial n} = \frac{\partial\psi}{\partial s}, \tag{8.2}$$

$$(1 - W^2)^{1/\gamma-1}W = \frac{\partial\varphi}{\partial s} + \frac{\partial\psi}{\partial n} \tag{8.3}$$

hold. The continuity equation becomes

$$\nabla^2\varphi = \left(1 + \frac{\gamma+1}{\gamma-1}\,W^2\right)(1 - W^2)^{(2-\gamma)/(\gamma-1)}q. \tag{8.4}$$

It is not advantageous to re-express the equation of motion and its consequence, Eq. (7.10), in terms of φ, ψ. An expression for ω can be found however. Our development is similar to that of Crocco and of Vazsonyi for the adiabatic case.

We first write (8.1) in component form

$$(1 - W^2)^{1/\gamma-1}u = \varphi_x + \psi_y\,, \tag{8.5}$$

$$(1 - W^2)^{1/\gamma-1}v = \varphi_y - \psi_x\,, \tag{8.6}$$

where $W = u\mathbf{i} + v\mathbf{j}$ and the subscripts indicate partial differentiation. We differentiate (8.5) with respect to y and let

$$(a^2/V_i^2) = c^2 = \frac{\gamma-1}{2}\,(1 - W^2). \tag{8.7}$$

Then,

$$(1 - W^2)^{1/\gamma-1}u_y = (1 - W^2)^{1/\gamma-1}\,\frac{Wu}{c^2}\,W_y + \varphi_{xy} + \psi_{yy}\,. \tag{8.8}$$

By symmetry also

$$(1 - W^2)^{1/\gamma-1}v_x = (1 - W^2)^{1/\gamma-1}\,\frac{Wv}{c^2}\,W_x + \varphi_{xy} - \psi_{xx} \tag{8.9}$$

and therefore

$$\begin{aligned} -(1 - W^2)^{1/\gamma-1}(v_x - u_y) &= -(1 - W^2)^{1/\gamma-1}\omega \\ &= (1 - W^2)^{1/\gamma-1}\,\frac{W}{c^2}\,(-vW_x + uW_y) + \nabla^2\psi. \end{aligned} \tag{8.10}$$

This equation is the same as the equation on p. 7 of Crocco's paper and therefore is of the same form whether the flow is diabatic or not. The terms W_x and W_y can be eliminated. From (8.5), (8.6)

$$(1 - W^2)^{2/\gamma-1}W^2 = \varphi_x^2 + \varphi_y^2 + \psi_x^2 + \psi_y^2 + 2\varphi_x\psi_y - 2\varphi_y\psi_x . \tag{8.11}$$

The gradient of this expression can be reduced to the form

$$W(1 - W^2)^{1/\gamma-1}\left(1 - \frac{W^2}{c^2}\right)\nabla W = u\nabla(\varphi_x + \psi_y) + v\nabla(\varphi_y - \psi_x). \tag{8.12}$$

Now taking the cross product with (8.1), we find

$$\begin{aligned} W\left(1 - \frac{W^2}{c^2}\right)(1 - W^2)^{1/\gamma-1}(uW_y - vW_x) = {} & -uv(\varphi_{xx} + \psi_{yy}) - v^2(\varphi_{xy} - \psi_{xx}) \\ & + u^2(\varphi_{xy} + \psi_{yy}) + uv(\varphi_{yy} - \psi_{xy}). \end{aligned} \tag{8.13}$$

Combination of (8.10) and (8.13) yields the desired formula for ω.

$$\begin{aligned} -(1 - W^2)^{1/\gamma-1}\omega = \left(1 - \frac{W^2}{c^2}\right)^{-1}\Bigg[& \frac{uv}{c^2}(\varphi_{yy} - \varphi_{xx}) + \frac{u^2 - v^2}{c^2}\varphi_{xy} \\ & + \left(1 - \frac{u^2}{c^2}\right)\psi_{xx} + \left(1 - \frac{v^2}{c^2}\right)\psi_{yy} - 2\frac{uv}{c^2}\psi_{xy}\Bigg] \end{aligned} \tag{8.14}$$

in which u, v, W, c are to be evaluated from (8.5), (8.6), (8.7).

This equation is still quasi-linear in φ and ψ. It reduces to Crocco's equation (10″) if $\varphi = 0$. For the incompressible case, $W \ll 1$, it becomes simply

$$\nabla^2\psi = -\omega \tag{8.15}$$

as in adiabatic flow whereas (8.4) becomes

$$\nabla^2\varphi = q_W \tag{8.16}$$

The symmetry in q_W , ω of Eqs. (7.16), (8.15), and (8.16) has been remarked upon previously [H-3].

When $q = 0$, φ may be taken to vanish and (8.14) then is a partial differential equation for ψ which changes type, as W increases through the value c, from elliptic to hyperbolic. When $\omega = 0$, it is appropriate to use φ_W and Eq. (5.4) gives similar behavior at $W = c = [(\gamma - 1)/(\gamma + 1)]^{1/2}$. In general, with neither q nor $\omega = 0$, φ and ψ are to satisfy (7.11)′, (8.4), and (8.14). Since there are three equations for the four dependent variables φ, ψ, q, ω, the system is underdetermined until one of them is specified independently.

9. The generation of vorticity by nonuniform heating. Let us suppose that in a field of compressible flow described in **W** language that ω_W is equal to zero upstream of a region R (see figure) and that only within R is q different from zero. Consider the curve N passing through R and intersecting all streamlines orthogonally. At A and B, outside R and on N, $q \doteq 0$ but within R, q is different from zero. Then the rate of variation of q along N, represented by $\partial q/\partial n$, must also be different from zero somewhere between A and B, say at C among other places. Along the streamline through C, ω must then be changing at C, according to (7.12) for, in general, the terms $2q\omega/W - \omega\partial \log p/\partial s + q\partial \log (1 - W^2)/\partial n$ will not cancel one another. Accordingly, on each streamline which passes through a point on N (or other such curves) where $\partial q/\partial n \neq$

0, $\partial\omega/\partial s$ will also not equal zero, and the value of ω on some streamlines issuing from R will differ from zero unless the q distribution is of such a character that $\int_R (\partial\omega/\partial s)\, ds = 0$ for every streamline passing through R.

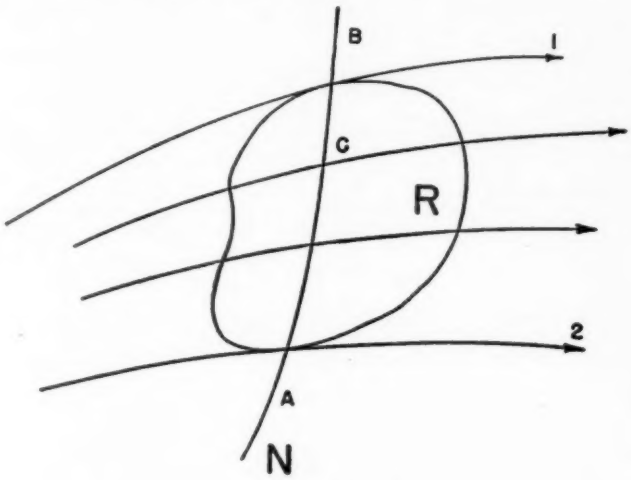

FIG. 1.

This situation also obtains in the incompressible case, as may be seen by referring to Eq. (7.16). When q is constant within R and changes suddenly at the boundaries of R to zero, then only along the streamlines (1) and (2) tangent to the boundary will ω change in the incompressible case. The flow which has passed through R will have been accelerated (or decelerated) leaving vortex lines (1) and (2) to effect the velocity jump to the flow outside (1) and (2). The heated region will therefore create a disturbance not unlike the wakes that have been studied in adiabatic flow.

IV. APPENDIX

10. Unit-vector formulas. Let $\mathbf{s}$ be defined as the unit vector in the direction of flow.

$$\mathbf{s} = \mathbf{W}/W = \mathbf{V}/V \tag{10.1}$$

and $\mathbf{n}$ be a second unit vector in the plane of the flow and normal to $\mathbf{s}$ such that

$$\mathbf{n} = \mathbf{k} \times \mathbf{s} \tag{10.2}$$

from which it follows that

$$\begin{aligned} \mathbf{k} &= \mathbf{s} \times \mathbf{n} \\ \mathbf{s} &= \mathbf{n} \times \mathbf{k} \end{aligned} \tag{10.3}$$

After the abbreviations

$$\mathbf{s}\cdot\nabla = \frac{\partial}{\partial s}, \qquad \mathbf{n}\cdot\nabla = \frac{\partial}{\partial n} \tag{10.4}$$

are introduced, it can be shown that

$$\mathbf{s} \times \nabla = \mathbf{k} \frac{\partial}{\partial n}, \tag{10.5}$$

$$\mathbf{n} \times \nabla = -\mathbf{k} \frac{\partial}{\partial s}, \tag{10.6}$$

$$\nabla \times (B\mathbf{n}) = \mathbf{k} \nabla \cdot (B\mathbf{s}), \tag{10.7}$$

where B is an arbitrary scalar function. These are the equations used in the development of (7.11).

It can also be shown that

$$\nabla \cdot \mathbf{n} = -\mathbf{k} \cdot \nabla \times \mathbf{s}, \tag{10.8}$$

$$\nabla \times \mathbf{n} = \mathbf{k} \nabla \cdot \mathbf{s}, \tag{10.9}$$

and these quantities can be re-expressed in terms of derivatives of θ, the angle of inclination of $\mathbf{s}$ to the x axis. Thus

$$\mathbf{s} = \mathbf{i} \cos \theta + \mathbf{j} \sin \theta, \tag{10.10}$$

$$\mathbf{n} = -\mathbf{i} \sin \theta + \mathbf{j} \cos \theta, \tag{10.11}$$

$$\nabla \cdot \mathbf{s} = -\theta_x \sin \theta + \theta_y \cos \theta = \mathbf{n} \cdot \nabla \theta \tag{10.12}$$

or

$$\nabla \cdot \mathbf{s} = \frac{\partial \theta}{\partial n}. \tag{10.13}$$

Similarly

$$\nabla \times \mathbf{s} = \mathbf{k} (\theta_x \cos \theta + \theta_y \sin \theta) = \mathbf{k}\mathbf{s} \cdot \nabla \theta$$

and

$$\nabla \times \mathbf{s} = \mathbf{k} \frac{\partial \theta}{\partial s}. \tag{10.14}$$

Concluding Remarks

Three representations have been studied of steady fields of flow of an inviscid compressible fluid that contains distributed heat sources. Of these representations, that afforded by the vector $\mathbf{W} = \mathbf{V}/V_t$, where $\mathbf{V}$ is the velocity vector and V_t is the (variable) limiting velocity appears to be the most convenient for rotational flows. Thus the intimate connection between vorticity $\omega = |\nabla \times \mathbf{W}|$ and the heating factor $q = Q/V_t^3$ (Q is the energy added per unit time and mass) is shown in W language by the equation for uniplanar flow

$$2W^{-1} q\omega + \omega \frac{\partial}{\partial s} \log (\omega/p) = q \frac{\partial}{\partial n} \log [q/(1 - W^2)]$$

in which $\partial/\partial s$ and $\partial/\partial n$ denote spatial differentiation parallel and perpendicular to streamlines and p is the (static) fluid pressure.

It has proved to be convenient to broaden the concept of irrotational flow in order to include diabatic flow fields in which $\mathbf{V}$, $\mathbf{W}$ or $\mathbf{M} = \mathbf{V}/a$ (a is the local velocity of sound)

is irrotational. Although the character of the corresponding heat source distributions is decidedly different in the three types of irrotational flow, the partial differential equations for the potential functions φ_s , φ_W and φ_M are all quasilinear and, except in the case of φ_s each contains an arbitrary function of the potential function, which is a consequence of the diabatic nature of the flow.

The question as to whether other representations than the three here studied may also be of interest in investigations of diabatic flow will be discussed in a subsequent paper describing the characterization of fields of diabatic flow.

References

B-1. H. Bateman, *The aerodynamics of reacting substances*, Proc. Nat. Acad. Sci. **25**, 388-391 (1939).

B-2. R. Becker, *Impact waves and detonation*, NACA TM Nos. 505, 506 [from Zeits. f. Phys. **8**, 321-362 (1922)].

C-1. L. Crocco, *Eine neue Stromfunktion für die Erforschung der Bewegung der Gase mit Rotation*, Zeits. angew. Math. Mech. **17**,, 1-7 (1937) [from Rend. Acad. Lincei (6) **23**, 1° sem., fasc. 2 (1936)].

H-1. B. L. Hicks, *Addition of heat to a compressible fluid in motion*, NACA ACR No. E5A29, Feb. 1945 (Wartime Report E-88).

H-2. ———, *Diabatic flow of a compressible fluid*, Phys. Rev. **69**, 135 (A), (1946).

H-3. ———, *Diabatic, rotational flow in two-dimensions*, Phys. Rev. **69**, 250 (A) (1946).

H-4. ———, P. E. Guenther and R. H. Wasserman, *New formulations of the equations for steady compressible flow*, Q. Appl. Math. **5**, 357-361 (1947).

H-5. ———, D. J. Montgomery and R. H. Wasserman, *On the one-dimensional theory of steady compressible fluid flow in ducts with friction and heat addition*, NACA ARR No. E6E22, June 1945 (TN No. 1336, 1947); J. Appl. Phys. **18**, 891-902 (1947).

H-6. ———, *On the characterization of fields of diabatic flow. Part I—General theory of steady diabatic flow*, Ballistic Research Laboratories Report No. 633, May 10, 1947 [see also Phys. Rev. **71**, 476 (A) (1947)], to appear in this Quarterly.

H-7. B. Haurwitz, *Dynamic meteorology*, McGraw-Hill, New York, 1941.

K-1. J. Kiebel, *Conditions for the dynamical possibility of motion of a compressible fluid with given addition of energy*, Rec. Geophys. (Leningrad) **5**, 3-46 (1932) (in Russian with French summary).

S-1. K. W. Sorg, *Supersonic flow in turbines and compressors*, J. Roy. Aero. Soc. **46**, 64-85 (1942) [from Forschung **10**, 270-285 (1939)].

T-1. W. Tollmien, *A vortex theorem for iso-energetic flow*, R.T.P. No. 1744 [from Luftfahrforschung **19**, 145-147 (1942)].

T-2. H. Tsien and L. Lees, *The Glauert-Prandtl approximation for subsonic flows of a compressible fluid*, J. Aero. Sci. **12**, 173-188 (1945).

V-1. Th. von Kármán, *Compressibility effects in aerodynamics*, J. Aero. Sci. **8**, 337-356 (1941).

V-2. A. Vazsonyi, *On rotational gas flows*, Quart. Appl. Math. **3**, 29-37 (1945).

A NEW TREATMENT OF THE LIFTING-LINE WING THEORY, WITH APPLICATIONS TO RIGID AND ELASTIC WINGS*

BY

W. R. SEARS

Cornell University

1. Introduction: the Prandtl airfoil theory. The approximate theory of three-dimensional airfoils known as the "lifting-line" theory was formulated in 1918 by L. Prandtl.[1] It has been in general use since then, and a number of different techniques have been developed for the numerical solution of the equations involved. Notable among these are the methods of Glauert,[2] Lotz,[3] and Multhopp.[4] The first two of these are based upon the use of trigonometric series, while the third involves solution of an integral equation by successive numerical approximations. In the present treatment, the Prandtl theory is formulated in terms of an integral equation of a classical type, and a practical solution in terms of its eigenfunctions is carried out. It is believed that certain advantages are offered by this procedure.

In the Prandtl theory the wing is assumed to be at rest in an incompressible fluid of infinite extent, whose velocity is uniform far upstream. The effects of viscosity are neglected, except as they are required to explain the presence of circulation about the wing. The wing is then replaced by a single vortex filament, or "lifting line", having at any spanwise station a strength equal to the circulation about a contour enclosing the wing at that station. There extends downstream from the lifting line a discontinuity surface, called the trailing-vortex sheet; hence the region is multiply connected. It is assumed that all the velocity components induced by the vortex filament and the vortex sheet are small in comparison with the undisturbed stream velocity; the vortex sheet may then be assumed to be approximately plane and to lie in the direction of the undisturbed velocity vector, at least for purposes of calculating certain induced velocity components at the lifting line.

The remaining approximation of the Prandtl theory is required to relate the circulation Γ about any section of the wing to the geometry of the wing at that section and the induced velocity at the same point on the lifting line. Let us select a coordinate system such that the distance along the span is denoted by y, as indicated in Fig. 1; let U denote the undisturbed stream velocity, $c(y)$ and $\alpha(y)$ the chord and angle of attack of the wing at y, and $w(y)$ the z component of the induced velocity at point $(0, y, 0)$ on the lifting line. The relation assumed by Prandtl, by analogy with the production of circulation about an airfoil in two-dimensional flow, is

$$\Gamma = \frac{1}{2} Ucm\left(\alpha - \frac{w}{U}\right), \tag{1}$$

where $m(y)$ is a number that is characteristic of the airfoil profile at y, namely the slope of the lift-coefficient curve for two-dimensional flow.

The most important practical problem of three-dimensional airfoil theory is the determination of $\Gamma(y)$ when the geometrical properties α, c, and m are prescribed functions of y; this requires the calculation of $w(y)$ in terms of $\Gamma(y)$. The strength of the vortex sheet, which according to the approximations stated above does not vary

*Received Nov. 28, 1947.

with x, is given by $-d\Gamma/dy$. Prandtl calculated w by integrating over the sheet; however, a treatment that is more convenient here is the one employed by Trefftz,[5] who considered conditions in a transverse plane far downstream where the effects of the lifting line are negligible. In such a plane the induced flow is steady and two-dimensional, and by symmetry the induced velocity components in the y and z directions are twice as great as at the corresponding points in the plane of the lifting line.

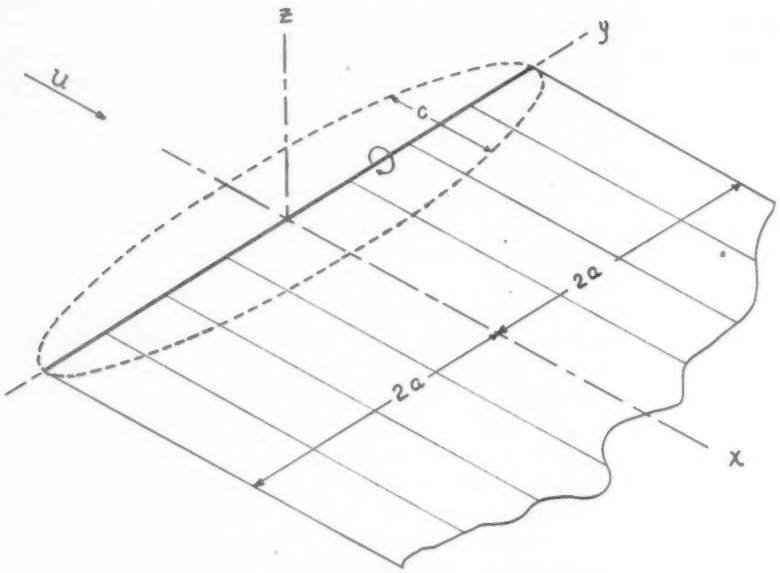

FIG. 1. Diagram showing notation.

Let $\varphi(y, z)$ be the velocity potential of the induced flow in Trefftz' plane. The jump in the value of φ at the discontinuity surface is seen to be equal to $\Gamma(y)$; i.e.

$$\begin{aligned}\Gamma(y) &= \varphi(y, +0) - \varphi(y, -0) \\ &= 2\varphi(y), \text{ say.}\end{aligned} \tag{2}$$

Hence the problem represented by Eq. (1) can be replaced by the following two-dimensional boundary-value problem:

$$\left.\begin{aligned} &\Delta\varphi \equiv \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} = 0 \\ &\text{(except on the slit } z = 0, \quad -2a \leqq y \leqq 2a\text{)}, \\ 2\varphi = &\quad \frac{1}{2}\,Ucm\left(\alpha + \frac{1}{2U}\frac{\partial\varphi}{\partial z}\right), \quad \text{for} \quad z = +0, \\ = &-\frac{1}{2}\,Ucm\left(\alpha + \frac{1}{2U}\frac{\partial\varphi}{\partial z}\right), \quad \text{for} \quad z = -0. \end{aligned}\right\} \tag{3}$$

It should be noted that, whereas the velocity potential has equal and opposite values on the two sides of the discontinuity surface, the z component of the velocity is continuous at this surface.

2. Formulation as an integral equation. It is convenient at this point to employ the transformation

$$y + iz = a(\zeta + \zeta^{-1})$$

by which the region exterior to the trace of the vortex sheet is mapped on the region outside the unit circle in the ζ plane. Here $4a$ denotes the span of the wing. If $\zeta \equiv re^{i\theta}$, the second and third of Eqs. (3) are replaced by the equations

$$\left.\begin{aligned} 2\varphi &= \frac{1}{2} Ucm\left(\alpha + \frac{1}{4a\, U \sin\theta}\frac{\partial\varphi}{\partial r}\right), \\ \varphi(\theta) &= -\varphi(2\pi - \theta), \end{aligned}\right\} \quad \text{for } r = 1, \qquad 0 \leqq \theta \leqq 2\pi$$

or, after rearrangement of terms,

$$\left.\begin{aligned} 4\frac{\sin\theta}{cm}\varphi - \frac{1}{4a}\frac{\partial\varphi}{\partial r} &= U\alpha \sin\theta, \\ \varphi(\theta) &= -\varphi(2\pi - \theta), \end{aligned}\right\} \quad \text{for } r = 1, \qquad 0 \leqq \theta \leqq 2\pi. \tag{4}$$

If we introduce the notations

$$\left.\begin{aligned} f(\theta) &\equiv \frac{(cm)_0}{cm}\sin\theta = f(2\pi - \theta), \\ g(\theta) &\equiv U(cm)_0\,\alpha \sin\theta = g(2\pi - \theta), \end{aligned}\right\} \quad 0 \leqq \theta \leqq 2\pi$$
$$l \equiv (cm)_0/4a > 0, \tag{5}$$

where $(cm)_0$ is the value of the product cm at a convenient reference station, say at $y = 0$, then we have the boundary-value problem in the form

$$\left.\begin{aligned} \Delta\varphi &= 0 \qquad \text{for} \qquad r > 1, \\ 4f\varphi - l\frac{\partial\varphi}{\partial r} &= g \qquad \text{for} \qquad r = 1, \\ \varphi(\theta) &= -\varphi(2\pi - \theta). \end{aligned}\right\} \tag{6}$$

Several procedures have been suggested for the numerical solution of this non-homogeneous boundary-value problem of the third kind. Lotz[3] used trigonometric expansions of the functions φ, f, and g; this leads to an infinite set of non-homogeneous linear equations for the Fourier coefficients of φ. This set is different for each choice of f and g, i.e. for each wing planform and angle-of-attack distribution. Schröder[6] and Gebelein[7] have set up the integral equation corresponding to the boundary-value problem, and Schröder has proposed a solution in terms of its eigenfunctions.

In the present treatment we formulate the integral equation by a somewhat different process, and carry out the practical solution in terms of its eigenfunctions. It appears

that this recourse to the underlying mathematical character of the problem leads to certain essential simplifications, since the eigenfunctions and eigenvalues are characteristic of a planform (f)—in fact, of a family of planforms—and do not depend upon the angle-of-attack distribution (g).

In view of the antisymmetrical character of the problem, Eq. (6) can immediately be replaced by an integral equation of the second kind by use of the Poisson equation for this case:

$$\varphi(\theta) = -\frac{1}{\pi}\int_0^{2\pi} \log\left|\csc\frac{\theta-\tau}{2}\right| \frac{\partial\varphi}{\partial r}\,d\tau$$

$$= -\frac{4}{\pi l}\int_0^{2\pi} \log\left|\csc\frac{\theta-\tau}{2}\right| f(\tau)\varphi(\tau)\,d\tau$$

$$+\frac{1}{\pi l}\int_0^{2\pi} \log\left|\csc\frac{\theta-\tau}{2}\right| g(\tau)\,d\tau. \tag{7}$$

If we multiply both sides of Eq. (7) by $[f(\theta)]^{1/2}$ and denote $f^{1/2}\,\varphi$ by Φ, we obtain a new integral equation involving a symmetrical kernel $K(\theta, \tau) = K(\tau, \theta) = \log|\csc(\theta-\tau)/2|\,[f(\theta)f(\tau)]^{1/2}$; i.e. the equation

$$\Phi(\theta) = \lambda\int_0^{2\pi} K(\theta, \tau)\Phi(\tau)\,d\tau + F(\theta), \tag{8}$$

where we have also written $-4/\pi l \equiv \lambda$ and

$$F(\theta) \equiv \frac{1}{\pi l}\int_0^{2\pi} K(\theta, \tau)\frac{g(\tau)}{[f(\tau)]^{1/2}}\,d\tau.$$

3. Solution of the integral equation. This integral equation [Eq. (8)] is of a classical type, having a real, symmetrical kernel of bounded integral-square, and there are available for its solution the theories of E. Schmidt, D. Hilbert, and others.* The solution of Eq. (8) is as follows:

(i)
$$\Phi(\theta) = F(\theta) + \lambda\left\{\frac{F_1}{\lambda_1-\lambda}\Phi_1(\theta) + \frac{F_2}{\lambda_2-\lambda}\Phi_2(\theta) + \cdots\right\}, \tag{9}$$

where $\lambda_1, \lambda_2, \cdots$ are the eigenvalues and $\Phi_1(\theta), \Phi_2(\theta), \cdots$ the corresponding normalized eigenfunctions of the homogeneous integral equation

$$\Phi_i(\theta) = \lambda_i\int_0^{2\pi} K(\theta, \tau)\Phi_i(\tau)\,d\tau; \qquad i = 1, 2, \cdots \tag{10}$$

and the definition of $F_1, F_2, \cdots$ is given by

$$F_i \equiv \int_0^{2\pi} F(\theta)\Phi_i(\theta)\,d\theta; \qquad i = 1, 2, \cdots. \tag{11}$$

Moreover,

(ii) the homogeneous equation [Eq. (10)] has infinitely many real eigenvalues, and, since the kernel $K(\theta, \tau)$ can be shown to be positive-definite,** all of its eigenvalues are

*See, for example, Reference 8, pp. 527-535, also Reference 9, pp. 104-118.

**Reference 9, p. 105.

positive;

(*iii*) the eigenfunctions of the homogeneous equation are real functions of θ in the interval $0 \leqq \theta \leqq 2\pi$, and constitute a complete system; they possess the property of orthogonality:

$$\int_0^{2\pi} \Phi_i(\theta)\Phi_j(\theta)\, d\theta = \delta_{ij}; \tag{12}$$

(*iv*) any arbitrary function $y(\theta)$ can be represented in the interval $0 \leqq \theta \leqq 2\pi$ by the series

$$y(\theta) = c_1\Phi_1(\theta) + c_2\Phi_2(\theta) + \cdots, \tag{13}$$

where

$$c_i \equiv \int_0^{2\pi} y(\theta)\Phi_i(\theta)\, d\theta; \qquad i = 1, 2, \cdots$$

and the series (13) is absolutely and uniformly convergent in the interval; provided that $y(\theta)$ can be represented in the form

$$y(\theta) = \int_0^{2\pi} K(\theta, \tau)z(\tau)\, d\tau. \tag{14}$$

The function $F(\theta)$ has been defined in the form (14); consequently it can be expanded in the form of equation (13):

$$F(\theta) = F_1\Phi_1(\theta) + F_2\Phi_2(\theta) + \cdots. \tag{15}$$

Thus a simpler form of Eq. (9) is obtained:

$$\Phi(\theta) = \frac{\lambda_1}{\lambda_1 - \lambda} F_1\Phi_1(\theta) + \frac{\lambda_2}{\lambda_2 - \lambda} F_2\Phi_2(\theta) + \cdots. \tag{16}$$

Moreover, if the definition of $F(\theta)$ is substituted into Eq. (11), one has

$$F_i = -\frac{\lambda}{4}\int_0^{2\pi}\int_0^{2\pi} K(\theta, \tau)\frac{g(\tau)}{[f(\tau)]^{1/2}}\Phi_i(\theta)\, d\tau\, d\theta$$

$$= -\frac{1}{4}\frac{\lambda}{\lambda_i}\int_0^{2\pi}\frac{g(\tau)}{[f(\tau)]^{1/2}}\Phi_i(\tau)\, d\tau; \qquad i = 1, 2, \cdots. \tag{17}$$

Hence, the solution of the lifting-line problem can be found immediately, in the form of a series expansion in eigenfunctions [Eq. (16)], where the coefficients are calculated from Eq. (17), provided, of course, that the eigenvalues and functions are known for the planform under consideration.

After discussing briefly the physical meaning of the eigenfunctions, we shall consider the problem of their numerical computation and present typical examples.

4. Physical interpretation. The preceding formulation has been largely mathematical. Although the theorems of the Hilbert-Schmidt theory tell us that the homogeneous Eq. (10) must have positive real eigenvalues, we have at present no insight into their physical significance. Returning to our original notation, we have

$$\lambda = -4/\pi l = -16a/\pi(cm)_0\,.$$

It is clear that positive eigenvalues correspond to negative values of the profile parameter m_0 , which is the reference value of the two-dimensional lift slope. Now, the homogeneous problem [Eq. (10)] amounts to the determination of the circulation distribution of the wing in question when its angle-of-attack distribution $\alpha(y)$ is identically zero. We see that this problem has non-trivial solutions only for discrete negative values of the lift slope.

Physically, this result is easily understandable. The untwisted wing at zero incidence can produce lift only when its induced velocities (downwash) act to *reinforce* the circulation, and this can occur only when m is negative! That this can occur only for certain discrete values also seems reasonable. The close analogy with the existence of self-reinforcing free vibrations of an elastic system at certain discrete natural frequencies is also obvious.

The eigenvalues are therefore the particular values of the parameter λ for which the wing planform in question, with no twist and zero incidence, has non-vanishing lift distributions. The eigenfunctions represent these distributions.

5. Numerical calculation of eigenfunctions. So far, we have assumed implicitly that the eigenvalues and -functions were known, so that the solutions represented in Eqs. (16) and (17) could be carried out. We shall now consider the practical problem of their numerical determination. There are numerous methods available for this calculation,* and a few of them have been tried out on this problem. One that is used in vibration problems has been adopted, but no assurance can be given that it is the most accurate or the easiest.

In this procedure, we resort to the approximate expression of the Φ_i in finite trigonometric series. The numerical calculation of each Φ_i is then analogous to that proposed by Lotz[2] for the inhomogeneous case, but can be simplified by virtue of its homogeneity.

Returning to our original notation, as in Eq. (7), we have for the homogeneous case

$$\varphi(\theta) = \lambda \int_0^{2\pi} \log \left| \csc \frac{\theta - \tau}{2} \right| \varphi(\tau) f(\tau)\, d\tau. \tag{18}$$

Let $\varphi(\theta) = \sum_1 c_n \sin n\theta$ and $f(\theta) = \sum_0 b_n \cos n\theta$. Then, using a known expression for the logarithmic kernel, we have

$$\sum_1 c_n \sin n\theta = \lambda \int_0^{2\pi} \sum_1^{\infty} \frac{\sin m\theta \sin m\tau}{m} \sum_1 c_j \sin j\theta \sum_0 b_k \cos k\theta \, d\theta. \tag{19}$$

Under suitable conditions of convergence, which are assumed here, this leads to a system of homogeneous linear equations for the c_n . The system can be solved approximately provided that the infinite series in Eq. (19) is terminated after the same number of terms as the other trigonometric series involved. This approximation leads to a matrix equation that can be solved numerically by a simple process of iterative multiplication:[11]

$$\frac{2}{\pi} \left\{ c \right\} = \lambda \left[k \right] \left\{ c \right\}, \tag{20}$$

where $\{c\}$ is the column matrix of c_n's, and $[k]$ is the square matrix

*See, for example, Reference 10 and the references listed there.

$$\begin{bmatrix} 2b_0 - b_2 & b_1 - b_3 & b_2 - b_4 & \cdots \\ (b_1 - b_3)/2 & (2b_0 - b_4)/2 & (b_1 - b_5)/2 & \cdots \\ (b_2 - b_4)/3 & (b_1 - b_5)/3 & (2b_0 - b_6)/3 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Numerical solution of Eq. (20) leads to an eigenvalue λ_1, and a corresponding matrix $\{c^{(1)}\}$, say, whose members are the Fourier constants of $\Phi_1/(f)^{1/2}$ or $\varphi_1(\theta)$, say $c_1^{(1)}$, $c_2^{(1)}$, $\cdots$. To determine the next eigenvalue and -function, the same numerical procedure is used, after eliminating the first set by one of the standard techniques, e.g. by use of the orthogonality relation

$$\sum m c_m^{(i)} c_m^{(j)} = 0 \qquad \text{if} \qquad i \neq j. \tag{21}$$

This procedure has been employed to calculate approximately the eigenvalues and -functions of a widely-used family of planforms, namely the family of "trapezoidal" wings, whose chord length $c(y)$ varies linearly from $y = 0$ to $y = 2a$. The results will be presented below.

6. Examples. *A. Elliptic planform.*

This case is represented by

$$cm = (cm)_0 \sin\theta$$

or

$$f = 1.$$

Hence

$$K(\theta, \tau) = \log \left| \csc \frac{\theta - \tau}{2} \right|,$$

and the eigenvalues and -functions for this case are*

$$\lambda_n = n/\pi; \qquad \Phi_n(\theta) = \pi^{-1/2} \sin n\theta.$$

The solution of an inhomogeneous lift-distribution problem for an elliptical planform is therefore given by

$$\Phi(\theta) = \varphi(\theta) = \Gamma/2 = 4aU \sum_1 A_n \sin n\theta,$$

$$A_n = \left(n + \frac{16a}{(cm)_0}\right)^{-1} B_n, \tag{22}$$

$$B_n \equiv \frac{2/\pi}{U(cm)_0} \int_0^\pi g(\tau) \sin n\tau \, d\tau = \frac{2}{\pi} \int_0^\pi \alpha \sin\tau \sin n\tau \, d\tau.$$

These formulas are well known.[12]

B. Trapezoidal planform. The planforms of greatest practical interest, in view of their common use in aeronautics, are those represented by

*Cf. Reference 9, p. 130.

$$cm = (cm)_0 \left\{ 1 - t \left| \frac{y}{2a} \right| \right\}; \qquad -2a \leqq y \leqq 2a$$

or

$$f = \frac{\sin \theta}{1 - t \mid \cos \theta \mid}; \qquad 0 \leqq \theta \leqq \pi.$$

where t is a constant.

TABLE I

EIGENVALUES FOR TRAPEZOIDAL WINGS

TAPER RATIO T	1	2	3	3.50	4.00
λ_1	.36726	.29665	.26553	.25537	.24722
λ_2	.8685	.61731	.51508	.48303	.45780
λ_3	1.3512	1.0031	.8480	.7982	.7585
λ_4	1.8286	1.3470	1.1305	1.0614	1.0065
λ_5	2.2950	1.7230	1.4504	1.3632	1.2944
λ_6	2.7572	2.0728	1.7400	1.6336	1.5496
λ_7	3.3209	2.4511	2.0900	1.9847	1.9008
λ_8	3.8448	2.8165	2.3789	2.2568	2.1625
λ_9	5.0487	4.4275	3.2631	2.9394	2.7021
λ_{10}	5.7616	5.2158	3.7958	3.4008	3.1093

TABLE II

Fourier coefficients $c_m^{(n)}$ of the functions $\varphi_n(\theta)$ (Odd values of n)

$c_m^{(n)}$	$T = 1$	$T = 2$	$T = 3$	$T = 3.5$	$T = 4.0$
$c_1^{(1)}$	+.59990	+.54354	+.51446	+.50375	+.49525
$c_3^{(1)}$	−.04933	−.01387	+.00644	+.01368	+.01974
$c_5^{(1)}$	+.00072	−.01100	−.01279	−.01278	−.01247
$c_7^{(1)}$	+.00053	+.00028	−.00066	−.00110	−.00147
$c_9^{(1)}$	+.00028	−.00094	−.00119	−.00132	−.00145
$c_1^{(3)}$	+.14955	+.02992	−.02691	−.04616	−.06183
$c_3^{(3)}$	+.60194	+.55224	+.52148	+.50782	+.49537
$c_5^{(3)}$	−.21574	−.12402	−.05528	−.03100	−.01099
$c_7^{(3)}$	+.03634	−.03257	−.04809	−.05003	−.05020
$c_9^{(3)}$	+.00082	+.00511	−.00068	−.00371	−.00635
$c_1^{(5)}$	+.07860	+.06207	+.05736	+.05714	+.05760
$c_3^{(5)}$	+.32339	+.15086	+.06048	+.02879	+.00268
$c_5^{(5)}$	+.47680	+.49449	+.50181	+.49618	+.48866
$c_7^{(5)}$	−.32639	−.24570	−.13417	−.09390	−.06070
$c_9^{(5)}$	+.10075	−.02524	−.07544	−.08500	−.08925
$c_1^{(7)}$	+.04596	+.02561	+.00600	+.00043	−.00365
$c_3^{(7)}$	+.20074	+.14254	+.11266	+.10890	+.10816
$c_5^{(7)}$	+.43372	+.21269	+.09220	+.05589	+.02895
$c_7^{(7)}$	+.39752	+.39032	+.45196	+.46896	+.48042
$c_9^{(7)}$	−.35774	−.34917	−.25385	−.20937	−.16527
$c_1^{(9)}$	+.00815	+.05221	+.03918	+.03555	+.03294
$c_3^{(9)}$	+.06094	+.16744	+.11292	+.09307	+.07569
$c_5^{(9)}$	+.22312	+.32499	+.24314	+.21699	+.19688
$c_7^{(9)}$	+.49177	+.44450	+.31287	+.25698	+.20307
$c_9^{(9)}$	+.58651	+.51959	+.49923	+.49608	+.49456

TABLE III

Fourier coefficients $c_m^{(n)}$ of the functions $\varphi_n(\theta)$ (Even values of n)

$c_m^{(n)}$	$T = 1$	$T = 2$	$T = 3$	$T = 3.5$	$T = 4.0$
$c_2^{(2)}$	+.62793	+.54543	+.50380	+.48897	+.47651
$c_4^{(2)}$	−.13722	−.07127	−.03155	−.01798	−.00694
$c_6^{(2)}$	+.01264	−.01852	−.02355	−.02384	−.02345
$c_8^{(2)}$	+.00163	−.00003	−.00268	−.00385	−.00481
$c_{10}^{(2)}$	+.00064	−.00099	−.00122	−.00147	−.00175
$c_2^{(4)}$	+.25463	+.11692	+.04780	+.02407	+.00467
$c_4^{(4)}$	+.54549	+.52247	+.50865	+.49960	+.49017
$c_6^{(4)}$	−.28016	−.19186	−.10477	−.07303	−.04686
$c_8^{(4)}$	+.06867	−.02812	−.05609	−.06100	−.06282
$c_{10}^{(4)}$	−.00381	+.00828	+.00044	−.00417	−.00836
$c_2^{(6)}$	+.14045	+.09593	+.07476	+.06899	+.06557
$c_4^{(6)}$	+.37005	+.19703	+.10285	+.06753	+.03825
$c_6^{(6)}$	+.40287	+.43699	+.47565	+.47671	+.47605
$c_8^{(6)}$	−.35825	−.30691	−.18484	−.13793	−.09959
$c_{10}^{(6)}$	+.13026	−.00235	−.07443	−.08973	−.09808
$c_2^{(8)}$	+.07570	+.06768	+.03750	+.03002	+.02498
$c_4^{(8)}$	+.24065	+.17700	+.12625	+.11791	+.11455
$c_6^{(8)}$	+.45828	+.24371	+.11728	+.07741	+.04796
$c_8^{(8)}$	+.35072	+.32604	+.41322	+.43762	+.45529
$c_{10}^{(8)}$	−.36860	−.38416	−.29392	−.25128	−.20898
$c_2^{(10)}$	+.00688	+.09938	+.07132	+.06221	+.05477
$c_4^{(10)}$	+.06776	+.22027	+.15621	+.13367	+.11401
$c_6^{(10)}$	+.24126	+.36526	+.27584	+.24644	+.22295
$c_8^{(10)}$	+.50300	+.46377	+.34179	+.29092	+.24157
$c_{10}^{(10)}$	+.58052	+.49813	+.47927	+.47762	+.47815

The approximate calculation described above has been applied to this case.

It is important to notice that f depends on the taper parameter t, but not on the aspect ratio. Consequently, the terms in Eq. (20) are independent of aspect ratio, and each set of eigenvalues and -functions can be used for an entire family of planforms having the same taper ratio.

The numerical results for this class of wings are summarized in Tables I–III and Figures 2–5. Figures 2 and 3 present the first ten eigenvalues as functions of the taper ratio T, which is defined as

$$T \equiv \frac{(cm)_0}{(cm)_{tip}} = \frac{1}{1 - t}.$$

In Figures 4 and 5 are shown a typical set of eigenfunctions Φ_n, namely the set corresponding to the case $T = 1$ (rectangular wing), for $n = 1$ to 10.

In Tables II and III are presented the values of the first ten Fourier coefficients $c_n^{(m)}$ of the functions $\varphi_m(\theta)$, for $m = 1$ to 10 and for several taper ratios between 1 and 4. It will be recalled that these functions are related to the normalized eigenfunctions by the relation

$$\Phi_m(\theta) = \varphi_m(\theta)[f(\theta)]^{1/2}.$$

These tables include all of the numerical data required to use the present method of solution with trapezoidal planforms within the range of values of T indicated.

7. Formulas for practical application (rigid wings). In this section we shall present

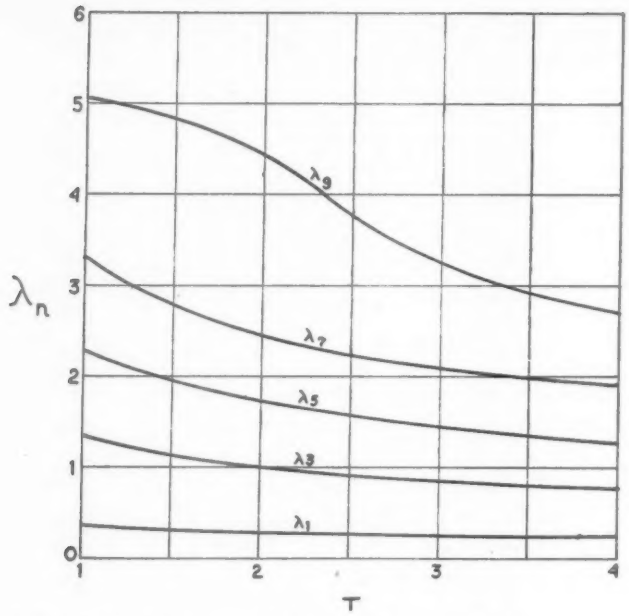

FIG. 2. Eigenvalues λ_n for trapezoidal wings of taper ratio T. Odd values of n.

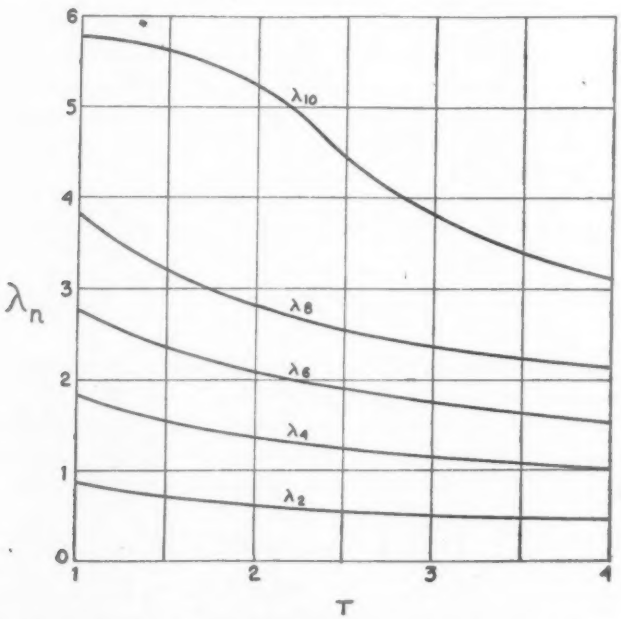

FIG. 3. Eigenvalues λ_n for trapezoidal wings of taper ratio T. Even values of n.

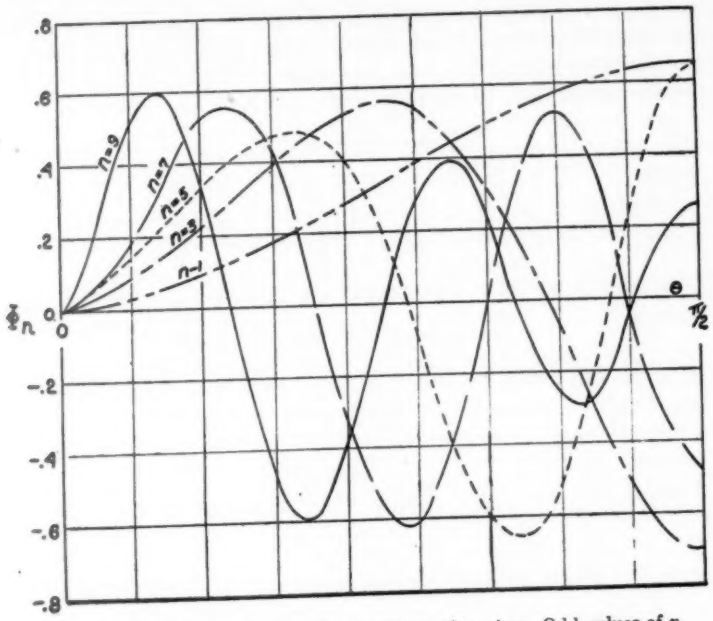

FIG. 4. The eigenfunctions $\Phi_n(\theta)$ for rectangular wings. Odd values of n.

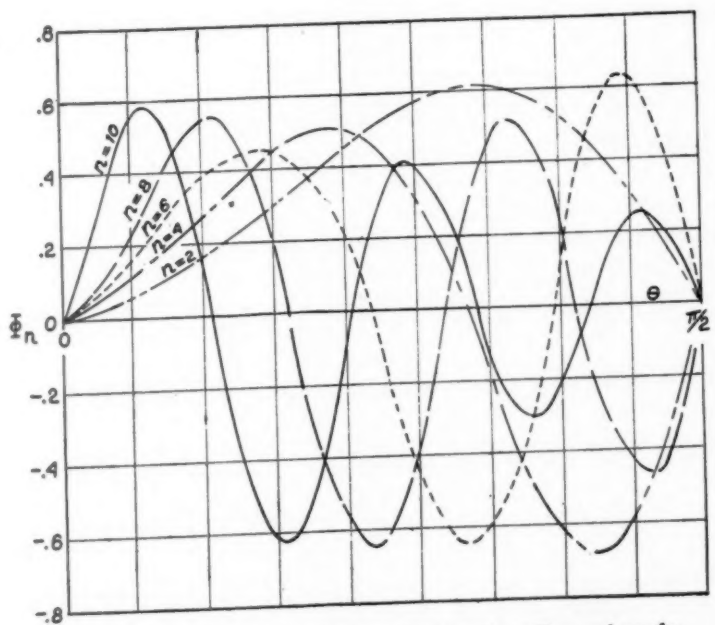

FIG. 5. The eigenfunctions $\Phi_n(\theta)$ for rectangular wings. Even values of n.

certain formulas that may be convenient in the practical application of the present theory to the classical problem of the Prandtl wing theory, i.e. to calculate the forces and the spanwise distribution of lift on a prescribed rigid wing.

It has already been found, in Eqs. (16) and (17), that

$$\Phi(\theta) = 4Ua\sum \mathfrak{A}_n\Phi_n(\theta), \tag{23}$$

where

$$\mathfrak{A}_n \equiv -\frac{1}{16Ua}\frac{\lambda}{\lambda_n - \lambda}\int_0^{2\pi} g(\tau)\varphi_n(\tau)\,d\tau.$$

Thus, the circulation distribution is given by

$$\Gamma = 2\varphi = 2\Phi/f^{1/2} = 8Ua\sum \mathfrak{A}_n\varphi_n(\theta). \tag{24}$$

When, as in the case of the trapezoidal wings above, the functions φ_n are given in terms of their Fourier constants, it is convenient to calculate the coefficients $\mathfrak{A}_n$ as follows:

$$\begin{aligned}\mathfrak{A}_n &= -\frac{1}{16Ua}\frac{\lambda}{\lambda_n - \lambda}\int_0^{2\pi} g(\tau)\sum_1 c_m^{(n)}\sin m\tau\,d\tau \\ &= -\frac{\pi}{16Ua}\frac{\lambda}{\lambda_n - \lambda}\sum_1 c_m^{(n)}B_m\,,\end{aligned} \tag{25}$$

where $B_m \equiv (1/\pi)\int_0^{2\pi} g(\tau)\sin m\tau\,d\tau$; i.e. the Fourier coefficient of the angle-of-attack function g.

Moreover, the usual sine series for Γ may be useful; it is commonly written as

$$\Gamma = 8Ua\sum_1 A_n \sin n\theta. \tag{26}$$

It is seen that the two "Fourier" coefficients, for expansion in sines and in eigenfunctions, respectively, are related as follows:

$$A_n = \sum_1 c_n^{(m)}\mathfrak{A}_m. \tag{27}$$

Hence, certain formulas of the Prandtl theory[2,3,5,12] can be transformed immediately to our new formulation:

$$\begin{aligned}C_L &\equiv \text{(Coefficient of total lift)} = \pi A\sum_1 c_1^{(m)}\mathfrak{A}_m\,, \\ C_l &\equiv \text{(Coefficient of total rolling moment)} = -\frac{\pi}{4}A\sum_1 c_2^{(m)}\mathfrak{A}_m\,, \\ C_D &\equiv \text{(Coefficient of induced drag)} = \pi A\sum_m\sum_n \mathfrak{A}_m\mathfrak{A}_n\sum_k kc_k^{(m)}c_k^{(n)} \\ &= \pi A\sum_1 \lambda_m\mathfrak{A}_m^2\,,\end{aligned} \tag{28}$$

where A denotes the aspect ratio, a constant.

Other useful expressions can be obtained by the well-known expedient of dividing the circulation distribution into so-called "basic" and "additional" parts, which are

independent of C_L and of $g(\theta)$, respectively. These formulas will not be developed here, since they are quite analogous to those employed in other methods of calculating spanwise load distributions.[13]

An additional simplification occurs in most practical cases because the planform is symmetrical about its center-line $y = 0$. This makes $f(\theta)$ symmetrical about $\theta = \pi/2$, and it is then found that φ_1 , φ_3 , φ_5 , $\cdots$ are symmetrical while φ_2 , φ_4 , $\cdots$ are antisymmetrical; thus half of the coefficients $c_m^{(n)}$ are identically zero. The trapezoidal planforms of the preceding section provide an example of this. In this situation it is clear that symmetrical and antisymmetrical lift-distribution problems are independent of one another; i.e., symmetrical $\alpha(y)$'s will produce only symmetrical $\Gamma(y)$'s, and antisymmetrical $\alpha(y)$'s will produce only antisymmetrical $\Gamma(y)$'s. Obviously, practical problems involving arbitrary $\alpha(y)$'s will usually be handled by considering the symmetrical and anti-symmetrical parts of $\alpha(y)$ separately.

In summary, it may be said that, for rigid wings, the use of eigenfunctions reduces every lift-distribution calculation to the same simplicity that ordinarily occurs only for the elliptical planform. However, since the eigenfunctions are not conveniently tabulated and are not as simple as trigonometric functions, this simplification is partly illusory, and for practical use one actually works with the Fourier constants $c_n^{(m)}$. Nevertheless, the simplification represented by Eqs. (25), (26), and (27) is very great; one calculates the span loading Γ by a straightforward process requiring only the Fourier coefficients of the angle-of-attack function g, besides tabulated quantities. No procedure of successive approximations is required.

8. Formulas for practical application (elastic wings). In the case of the elastic wing, whose ultimate geometrical configuration is determined by the airload distribution, it appears that the present theory may offer rather important possibilities. At present, calculations for such wings are either made without attention to the aerodynamical principles of the Prandtl theory—i.e. by neglecting the effects of the trailing vortex sheet entirely—or by assuming special and simple geometrical and elastic properties, which may be completely inapplicable to wings encountered in practical engineering problems. Another method is to perform a process of successive approximations, estimating the elastic deflection from the rigid-wing airloads, repeating the rigid-wing calculation for the deflected wing, and so forth. It is clear that this process must fail at each of the so-called divergence speeds, where the elastic restoring force in some mode of deflection becomes insufficient to resist the air forces. It is also likely that the convergence of the process becomes poor as such a divergence situation is approached, and that this will make it impossible either to predict the divergence speed or to calculate air loads with satisfactory accuracy near such a speed.

To attack this problem, we first note explicitly that the load distribution $\rho U\Gamma(y)$* is linearly related to the angle-of-attack distribution $\alpha(y)$ or $g(\theta)$, according to the Prandtl theory considered here. Let us denote by $\mathcal{P}[g]$ the circulation distribution of a given wing at the speed U when the angle-of-attack function is $g(\theta)$. Thus $\mathcal{P}[g]$ is the "Prandtl functional" calculated in the preceding paragraphs, i.e.

$$\Gamma = \mathcal{P}[g]$$

and it is linear in g; i.e.

* ρ denotes the fluid density.

$$\mathcal{O}[g_1 + g_2] = \mathcal{O}[g_1] + \mathcal{O}[g_2],$$
$$\mathcal{O}[kg] = k\mathcal{O}[g]. \tag{29}$$

Now if the subscripts r and e denote, respectively, the contributions due to the built-in angle of attack, $\alpha_r(y)$, and the elastic deflection under load $\alpha_e(y)$, we have

$$\alpha(y) = \alpha_r(y) + \alpha_e(y),$$
$$g(\theta) = g_r(\theta) + g_e(\theta), \tag{30}$$
$$\Gamma = \mathcal{O}[g] = \mathcal{O}[g_r] + \mathcal{O}[g_e].$$

Moreover, the elastic deflection (twist) is determined by a linear functional involving the load distribution $\rho U\Gamma(y)$. Ordinarily, it is satisfactory to calculate this deflection by means of the torsional influence function $F(y, \eta)$, which denotes the angular twist at y due to a concentrated unit torque at η. This function can be calculated or it can be determined experimentally; for practical airplane wings it is often a very irregular function. It is used to calculate the elastic twist as follows:

$$\alpha_e(y) = \int_0^{2a} F(y, \eta)\frac{d\mathfrak{I}}{d\eta}\,d\eta,$$

where $\mathfrak{I}(y)$ is the torque about the elastic axis at any spanwise station y:

$$\frac{d\mathfrak{I}}{dy} = M + \rho U\Gamma e, \tag{31}$$

where $M = M(y) \equiv$ aerodynamic moment about aerodynamic center at y
$= (\tfrac{1}{2})\rho U^2c^2c_m(y)$,
$c_m(y) \equiv$ moment coefficient of the profile at y (a given function),
$e = e(y) \equiv$ distance between aerodynamic center and elastic axis at y (a given function).

Thus

$$\alpha_e(y) = \frac{1}{2}\rho U^2\left\{\int_0^{2a} F(y, \eta)c^2c_m\,d\eta + 2\int_0^{2a} F(y, \eta)e\mathcal{O}[g/U]\,d\eta\right\}$$

or

$$g_e/U = (cm)_0\alpha_e \sin\theta$$
$$= \frac{1}{2}\rho U^2(cm)_0 \sin\theta\left\{\int_0^{2a} F(y, \eta)c^2c_m\,d\eta\right.$$
$$\left. + 2\int_0^{2a} F(y, \eta)\left(\mathcal{O}\left[\frac{g_r}{U}\right] + \mathcal{O}\left[\frac{g_e}{U}\right]\right)e\,d\eta\right\}. \tag{32}$$

It should be noticed that there are two terms on the right-hand side of Eq. (32) involving only known functions and a third involving the elastic deflection. Equation (32) is a linear integral equation of second kind, but involves a very complicated kernel. Its solution might be found in terms of the eigenvalues and -functions of the homo-

geneous case, i.e. the case $c_m = g_r = 0$. These eigenvalues would represent the values of $\frac{1}{2}\rho U^2(cm)_0$ for which wing deflection would occur with no initial moment or angle of attack; viz., the divergence speeds.

For practical calculation, using the present technique of handling the Prandtl problem, we make use of a symbolism that is equivalent to Eqs. (6), (10), or (18):*

$$2\varphi_n(\theta) = \mathcal{P}\left[4\,\frac{\lambda - \lambda_n}{\lambda} f\varphi_n\right]$$

or (33)

$$\mathcal{P}[f\varphi_n] = \frac{\lambda/2}{\lambda - \lambda_n}\varphi_n .$$

where $f^{1/2}\varphi_n$ is the normalized eigenfunction, as before. This shows, incidentally, that the unique property of the eigenfunctions can alternatively be expressed thus: $f\varphi_n$ is the angle-of-attack distribution that produces, on the given wing, a circulation distribution proportional to φ_n.

Therefore, if we expand g_r/U and g_e/U in series of φ_n's, we can immediately operate term-by-term with the Prandtl functional symbol $\mathcal{P}$; i.e. if

$$\frac{g_r}{U} = f(\theta)\sum b_i^{(r)}\varphi_i(\theta) \qquad \text{and} \qquad \frac{g_e}{U} = f(\theta)\sum b_i^{(e)}\varphi_i(\theta),$$

then

$$\mathcal{P}[g/U] = \sum (b_i^{(r)} + b_i^{(e)})\mathcal{P}[f\varphi_i] = \sum (b_i^{(r)} + b_i^{(e)})\frac{\lambda/2}{\lambda - \lambda_i}\varphi_i . \tag{34}$$

But according to Eqs. (13),

$$\begin{aligned} b_n^{(e)} &= \int_0^{2\pi} \varphi_n(\theta)\frac{g_e}{U}\,d\theta \\ &= \frac{1}{2}\rho U^2(cm)_0\left\{\int_0^{2\pi} \varphi_n(\theta)\sin\theta \int_0^{2a} F(y,\eta)c^2c_m\,d\eta\,d\theta\right. \\ &\quad \left. + \sum_i (b_i^{(r)} + b_i^{(e)})\frac{\lambda}{\lambda - \lambda_i}\int_0^{2\pi} \varphi_n(\theta)\sin\theta \int_0^{2a} F(y,\eta)c\varphi_i\,d\eta\,d\theta\right\} \end{aligned} \tag{35}$$

(for $n = 1, 2, \cdots$).
This equation has the form

$$b_n^{(e)} = \frac{1}{2}\rho U^2(cm)_0\left\{C_n + D_n + \sum_i E_{ni}b_i^{(e)}\right\}. \tag{36}$$

---

*For, according to equation (6), the symbolism $2\varphi = \mathcal{P}[g]$ means

$$4f\varphi + (4/\pi\lambda)\partial\varphi/\partial r = g$$

in any case. But for the eigenfunctions, or rather the related φ_n,

$$4f\varphi_n + (4/\pi\lambda_n)\partial\varphi_n/\partial r = 0;$$

hence

$$4f\varphi_n + (4/\pi\lambda)\partial\varphi_n/\partial r = 4(1 - \lambda_n/\lambda)f\varphi_n .$$

where the terms have the following physical meanings:

C_n = nth "Fourier" coefficient, for expansion in $\varphi_n(\theta)$'s, corresponding to the twist due to the aerodynamic-center moments $c_m(y)$, at unit dynamic pressure.

$D_n \equiv \sum E_{ni} b_i^{(r)}$ = nth "Fourier" coefficient corresponding to the twist due to the rigid-wing airload distribution, at unit dynamic pressure.

$\sum_i E_{ni} b_i^{(e)}$ = nth "Fourier" coefficient corresponding to the twist due to the airload resulting from elastic twist, at unit dynamic pressure.

Moreover, E_{ni} is equal to $[\lambda/(\lambda - \lambda_i)]F_{ni}$, where F_{ni} is the nth "Fourier" coefficient for the series expansion of the twist produced by the circulation distribution φ_i, at unit dynamic pressure.

It is clear that the elastic-wing problem is solved when the $b_n^{(e)}$ are determined from Eqs. (36), since they define the only unknown function, the elastic deflection under load. It is seen that the essential calculation required for the elastic wing is the determination of a set of F_{ni}'s, i.e., one must calculate the elastic deflection under a loading proportional to each φ_i for a sufficient number of values of i, and must then determine the "Fourier" coefficients of these deflections. To do this, it is most convenient to use the formula

$$F_{ni} = \sum_m c_m^{(n)} \int_0^{2\pi} \Theta_i(\theta) \sin\theta \sin m\theta \, d\theta, \tag{37}$$

where

$$\Theta_i(\theta) \equiv \int_0^{2a} F(y, \eta) e\varphi_i \, d\eta.$$

Equations (36) are most conveniently solved by standard matrix methods.[14] Of the several techniques available, the partial-fraction type of expansion* gives the most engineering information, since the latent roots, which appear as simple poles in the expansion, correspond to the divergence speeds of the wing, as has been mentioned above.

Once again it will be found that, for a symmetrical planform, the symmetrical and antisymmetrical distributions are independent of one another. Thus a set of odd-numbered and a set of even-numbered latent roots appear, corresponding respectively to divergence in symmetrical and antisymmetrical modes, and so forth.

In summary, the following may be said regarding the present treatment of the problem of the elastic wing. The integral equation of the problem, Eq. (32), has been put into matrix form, and an essential feature of the treatment is the proposal that matrix techniques be used for numerical solution. Moreover, in making this formulation, essential use has been made of the knowledge of a complete set of functions, the eigenfunctions of the planform, for which the Prandtl functional can be written down immediately.

9. Acknowledgements. The use of the eigenfunctions of the Prandtl equation, as well as the physical interpretation of these functions, was suggested in 1939 by Professor Theodore von Kármán. Following his suggestion, the present author worked out the details while he was a staff member of the California Institute of Technology. The eigenvalues and -functions for trapezoidal planforms were calculated by H. Fischer in

*See, for example, reference 14, pp. 78, 134-141.

1940 and later by H. T. Nagamatsu, both of the California Institute. This method was used extensively by the engineering staff of Northrop Aircraft, Inc., and the application to elastic wings was worked out at Northrop by the author and H. R. Lawrence in 1943–44. The formulation in terms of matrices and the application of matrix techniques in practical computations are largely due to Mr. Lawrence.

Incidentally, it may be mentioned that Mr. Lawrence's work has progressed somewhat farther than the material presented here. He has extended the calculations, in practical form, to problems of aileron reversal, wing deflection and divergence in accelerated flight, and other similar cases. It is expected that his results along these lines will be published soon.

References

1. L. Prandtl, *Tragflügeltheorie*, Nachrichten d. K. Gesellschaft d. Wiss. zu Göttingen, Math.-phys. Klasse **1918**, 451-477. Reprinted in *Vier Abhandlungen zur Hydrodynamik und Aerodynamik*, Göttingen, 1927, or Edwards Bros., 1943.

2. H. Glauert, *Elements of the aerofoil and airscrew theory*, Cambridge, 1926.

3. I. Lotz, *Berechnung der Auftriebsverteilung beliebig geformter Flügel*, Zeitschr. Flugtechn. Motorluftschiffahrt **22**, 189-195 (1931).

4. H. Multhopp, *Die Berechnung der Auftriebsverteilung von Tragflügeln*, Luftfahrtforschung **15**, 153-169 (1938).

5. E. Trefftz, *Prandtlsche Tragflächen- und Propellertheorie*, Zeitschr. angew. Math. Mech. **1**, 206-218 (1921).

6. K. Schröder, *Die Prandtlsche Integro-differentialgleichung*, Abh. Preussischen Akad. Wiss. **1939**, Nr. 16 (35 pp.).

7. H. Gebelein, *Über die Integralgleichung der Prandtlschen Tragflächentheorie*, Ing.-Arch. **7**, 297-325 (1936).

8. P. Frank and R. von Mises, *Die Differential- und Integralgleichungen der Mechanik und Physik*, (2nd edition) Vieweg, Braunschweig, 1930, or Rosenberg, New York, 1943.

9. R. Courant and D. Hilbert, *Methoden der Mathematischen Physik I* (2nd edition) Springer, Berlin, 1931, or Interscience, New York, 1943.

10. L. Collatz, *Genäherte Berechnung von Eigenwerten*, Zeitschr. angew. Math. Mech. **19**, 224-249; 297-318 (1939).

11. Th. von Kármán and M. A. Biot, *Mathematical methods in engineering* (Chapter V), McGraw Hill, New York, 1940.

12. M. Munk, *Fluid dynamics for aircraft designers*, Ronald, New York, 1929, pp. 99-102.

13. R. F. Anderson, *Determination of the characteristics of tapered wings*, N. A. C. A. Technical Report No. 572 (1936).

14. R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary matrices and some applications to dynamics and differential equations*, Cambridge University Press, Cambridge, 1938.

ZONAL COMBUSTION IN TUBES[1]

BY

S. A. SCHAAF

University of California, Berkeley

1. Introduction. Theoretical work on combustion of gases has been chiefly concerned with two types of approximations: (*i*) "constant volume" or uniform state combustion,[2] where it is assumed that the state of a burning gas in a chamber is the same at all parts of the chamber at any given time; (*ii*) "detonation front" and "deflagration front" combustion,[3] where it is assumed that combustion takes place instantaneously in an infinitesimally thin front which advances with a finite velocity through the gas. We consider here a third type of approximation where it is assumed that combustion extends over a definite zone or region, which moves with time through the gas. This model is suggested by the appearance of turbulent combustion in tubes. We confine our attention to effectively one-dimensional phenomena, in particular to combustion in a half-open tube. A basic assumption is that at any time the effective pressure, velocity, etc. of the gas in a given cross-section of the tube can be characterized by single quantities. The important physical parameters, which will depend on the degree of turbulence in the tube, the strength of the combustible mixture, etc., are the rate with which energy is released during combustion, the velocity with which the burning region progresses into the fresh gas and the duration of burning of a given layer of gas. Explicit results are obtained good to first order terms in a perturbation expansion in a parameter proportional to the average rate of energy release.

2. Formulation. Let us consider a tube of constant cross-section and length L, (see Fig. 1) initially filled with a combustible gas at rest and at uniform normal temperature and pressure, open at one end and closed at the other. The combustible gas is ignited at the closed end, and the flame front travels down the tube followed by a zone of burning gas. We denote by x the distance from the closed end of the tube, by t the time, by a the initial position (x) of a given layer of gas, by p the pressure and by v the specific volume (reciprocal of the density). We consider the equations of the system

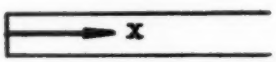

FIG. 1. Half-open tube filled with combustible gas.

in Lagrangian coordinates, so that a and t are the independent variables and $x = x(a, t)$, $p = p(a, t)$ $v = v(a, t)$. Assuming that at time $t = 0$ the gas in the tube is in a uniform state with $p = p_0$, $v = v_0$ and $x = a$, we have the equations of continuity and momentum respectively

[1]Received Sept. 5, 1947. This paper uses a type of formulation which was developed for the study of jet devices at the Institute of Mathematics and Mechanics of New York University by J. K. L. MacDonald and the author under a Navy contract. See *A gas dynamical formulation for waves and combustion in pulse-jets*, Applied Mathematics Group, New York University, Report 151.

[2]See, for example, Jennings and Obert, *Internal combustion engines*, International Textbook Co., 1944.

[3]See, for example, R. Becker, *Stosswelle und Detonation*, Z. Physik, **8**, 1922.

$$v = v_0 \frac{\partial x}{\partial a} \tag{1}$$

$$\frac{\partial^2 x}{\partial t^2} = -v_0 \frac{\partial p}{\partial a} \tag{2}$$

For the energy equation, we consider a given layer of gas in which energy of amount E per unit mass per unit time is being liberated by the combustion process. This energy will go partly into expansion of the layer and partly into raising the temperature.[4] We thus have

$$E = p \frac{\partial v}{\partial t} + \frac{\partial}{\partial t}(c_v \theta),$$

as well as the gas law

$$pv = (c_p - c_v)\theta,$$

where θ denotes the temperature, and c_v and c_p are the specific heats at constant volume and pressure respectively. Eliminating the temperature we obtain the energy equation

$$(\gamma - 1)E = \gamma p \frac{\partial v}{\partial t} + v \frac{\partial p}{\partial t}, \tag{3}$$

where γ is the adiabatic exponent, $\gamma = c_p/c_v$.

For the zonal type of combustion we are considering, the quantity E will at any time be zero in the non-burning parts of the tube and have non-zero values depending on the time, temperature, turbulence, gas mixture, etc., in the burning part. Let t_0 be the average duration of combustion in a given layer of gas, and ΔE be the average total energy released per unit mass by the combustion; then the average value of E in the burning region will be $E_0 = \Delta E/t_0$. The burning region will advance into the fresh gas with a certain velocity f relative to the gas, depending on the temperature, turbulence, etc. Let f_0 be the average value of f. Our principal idealization then is to replace the functions f and E by their average values, so that $f = f_0$, and $E = E_0$ in the burning region and $E = 0$ in the non-burning region,[5] i.e. we assume that

$$E = \begin{cases} E_0 & \text{for} \quad f_0 t > a > f_0(t - t_0) \quad \text{or} \quad 0 \\ 0 & \text{for} \quad a > f_0 t;\ a < f_0(t - t_0). \end{cases} \tag{4}$$

The differential system (1), (2) and (3) is then completely specified; we wish to obtain a solution of it subject to the following initial and boundary conditions,

[4]The complete energy balance of course includes terms accounting for heat flow and dissipation, which we consider as included in the E, as well as terms accounting for the kinetic energy of mass motion and work done by the pressure gradient; these latter two cancel out by virtue of the momentum equation (2), so that the energy equation as written is complete.

[5]It is not necessary to treat this as an idealization since these average values could be taken as only the first approximations to E and f in the perturbation process considered in section 3. The results to first order terms are, of course, the same; it seems preferable for clarity and brevity of presentation to regard this form of E and f as idealizations.

$$\text{at} \quad t = 0; \quad p = p_0, \quad v = v_0, \quad x = a$$

$$\frac{\partial p}{\partial t} = \frac{\partial v}{\partial t} = \frac{\partial x}{\partial t} = 0 \tag{5}$$

$$\text{at} \quad x = 0 : \frac{\partial x}{\partial t} = 0 \qquad \text{(closed end)}$$

$$\text{at} \quad x = L : p = p_0 \qquad \text{(open end)}$$

It is convenient to introduce the dimensionless quantities

$$V = v/v_0, \quad P = p/p_0, \quad X = x/L, \quad A = a/L,$$

$$T = ct/L, \quad T_0 = ct_0/L, \quad \lambda = f_0/c,$$

$$\epsilon = \gamma(\gamma - 1) E_0 L/c^3,$$

$$I = \begin{cases} 1 & \text{for} \quad \lambda T > A > \lambda(T - T_0), \\ 0 & \text{for} \quad A > \lambda T \quad \text{or} \quad A < \lambda(T - T_0), \end{cases}$$

where c is the initial sound speed, $c^2 = \gamma p_0 v_0$. The dimensionless form of (1), (2), (3) is then

$$V = \frac{\partial X}{\partial A}, \tag{1'}$$

$$\gamma \frac{\partial^2 X}{\partial T^2} = -\frac{\partial P}{\partial A}, \tag{2'}$$

$$\gamma P \frac{\partial V}{\partial T} + V \frac{\partial P}{\partial T} = I\epsilon. \tag{3'}$$

The initial and boundary conditions become

$$\text{at } T = 0 : P = V = 1, \quad X = A,$$

$$\frac{\partial P}{\partial T} = \frac{\partial V}{\partial T} = \frac{\partial X}{\partial T} = 0, \tag{5'}$$

$$\text{at } X = 0 : \frac{\partial X}{\partial T} = 0,$$

$$\text{at } X = 1 : P = 1.$$

3. Perturbation solution. The non-linear system (1'), (2'), (3') is not easily solvable in exact form, but approximate solutions can be obtained valid for "small" values of

the parameter ϵ, if we assume that a power series development of P, V and X in terms of ϵ is possible, so that

$$P = 1 + \epsilon P'(A,T) + \cdots$$

$$V = 1 + \epsilon V'(A,T) + \cdots$$

$$X = A + \epsilon X'(A,T) + \cdots$$

Substituting into (1'), (2') and (3'), and ignoring powers of ϵ higher than the first, we obtain a linear system

$$V' = \frac{\partial X'}{\partial A}, \tag{i}$$

$$\gamma \frac{\partial^2 X'}{\partial T^2} = -\frac{\partial P'}{\partial A}, \tag{ii}$$

$$\gamma \frac{\partial V'}{\partial T} + \frac{\partial P'}{\partial T} = I. \tag{iii}$$

For boundary and initial conditions, we have

$$\text{at } T = 0 : P' = V' = X' = \frac{\partial P'}{\partial T} = \frac{\partial V'}{\partial T} = \frac{\partial X'}{\partial T} = 0,$$

$$\text{at } A = 0 : \frac{\partial X'}{\partial T} = 0,$$

$$\text{at } A = 1 : P' = 0.$$

The independent variable A may be used instead of the dependent variable X in the end-point boundary conditions since $X = A$ to first order terms in the perturbation parameter ϵ.

The linear system (*i*), (*ii*), (*iii*) may be most easily solved by introducing a generating function $W(A, T)$ such that

$$\frac{\partial W}{\partial A} = 2\gamma(1 - \lambda^2)\frac{\partial X'}{\partial T}; \qquad \frac{\partial W}{\partial T} = 2(1 - \lambda^2)P'.$$

Then (*ii*) is solved identically. Eliminating V', P' and X', we obtain a wave equation for W,

$$\frac{\partial^2 W}{\partial A^2} - \frac{\partial^2 W}{\partial T^2} = 2(1 - \lambda^2)I. \tag{6}$$

The general solution of (6) will be of the form

$$W = F(A + T) + G(A - T) + (A - \lambda T)^2 I, \tag{7}$$

where the functions F and G must be determined so as to satisfy the boundary conditions

$$\text{at } A = 0, \qquad \frac{\partial W}{\partial A} = 0,$$

$$\text{at } A = 1, \qquad \frac{\partial W}{\partial T} = 0.$$

The initial conditions are that W and its first partial derivatives be zero at $t = 0$. We also require that W and its first partial derivatives be continuous for $t > 0$. In Fig. 2

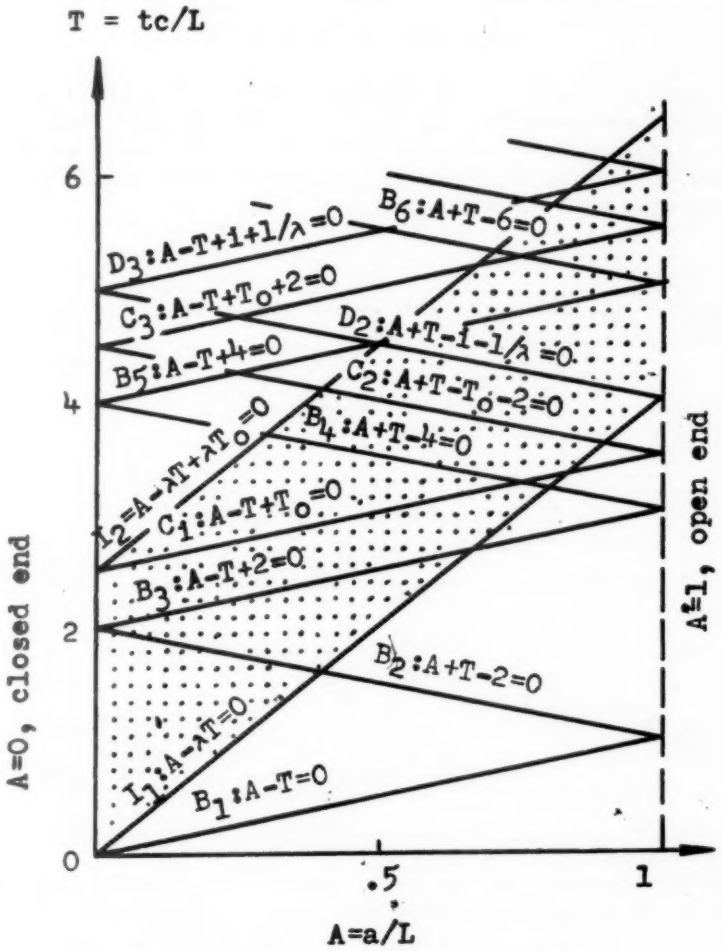

FIG. 2. Characteristic lines and burning zone in (A, T) plane.

we indicate the region of interest in the (A, T) plane, together with the important characteristic lines of the form $A \pm T = \text{const.}$; the shaded region corresponds to the burning region. There are three independent sets of characteristic lines, labeled with

the letters B, C and D respectively. It may be easily verified that the function W which satisfies all the required conditions is given by

$$\begin{aligned} W = {} & B_1^*(A - T)^2 + B_2^*(A + T - 2)^2 + B_3^*(A - T + 2)^2 + \cdots \\ & + I_1^*(A - \lambda T)^2 + I_2^*(A - \lambda T + \lambda T_0)^2 \\ & + C_1^*(A - T + T_0)^2 + C_2^*(A + T - 2 - T_0)^2 + \cdots \\ & + D_2^*(A + T - 1 - 1/\lambda)^2 + \cdots , \end{aligned} \tag{8}$$

where the quantities B_1^*, C_1^*, etc. are defined to be zero for all points (A, T) below the corresponding characteristic lines and equal to certain constants above these lines (see Fig. 2). The constants are determined so that the boundary conditions are satisfied. For example: in the region in the (A, T) plane near the closed end and just inside the burning region (see Fig. 2), W is given by

$$W = B_1(A - T)^2 + I_1(A - \lambda T)^2.$$

According to (7) and the definition of I, we have $I_1 = 1$. Then B_1 is to be determined so as to satisfy the boundary condition

$$\frac{\partial W}{\partial A} = 0 \qquad \text{at} \qquad A = 0.$$

We have

$$\left.\frac{\partial W}{\partial A}\right|_{A=0} = 2B_1(-T) + 2(-\lambda T) = 0,$$

and hence $B_1 = -\lambda$. The other constants are determined in a similar fashion. The result is

$$\begin{aligned} -\lambda = B_1 = -B_2 = -B_3 = B_4 = B_5 = -B_6 = \cdots , \\ \lambda = C_1 = -C_2 = -C_3 = \cdots , \\ \lambda^2 \quad = -D_2 = -D_3 = \cdots , \\ I_1 = 1, \qquad I_2 = -1. \end{aligned} \tag{9}$$

These constants completely specify the function W. The quantities P, X and $M = \partial X/\partial T = \partial x/c\partial t$ are then given in terms of W by,

$$P = 1 - \frac{\epsilon}{2(1 - \lambda^2)} \frac{\partial W}{\partial T} \tag{10}$$

$$X = A + \frac{\epsilon}{2\gamma(1 - \lambda^2)} \int_0^T \frac{\partial W}{\partial A}\, dT, \tag{11}$$

$$M = \frac{\epsilon}{2\gamma(1-\lambda^2)}\frac{\partial W}{\partial A}. \tag{12}$$

4. Particular results. Special cases of (10), (11) and (12) which are of particular interest are $P(0, T)$, $X(\lambda T, T)$ and $M(1, T)$ which give respectively the pressure at the closed end of the tube, the position of the flame front and the velocity of gas flow at the open end as functions of the time. These are given by

$$P = 1 + \frac{2\lambda\epsilon}{1-\lambda^2}\begin{cases} (1-\lambda)T, & T > 0, \\ -(T-2), & T > 2, \\ +(T-4), & T > 4, \\ -(T-6), & T > 6, \\ \cdots\cdots \\ -(1-\lambda)(T-T_0), & T > T_0, \\ +(T-T_0-2), & T > T_0+2, \\ -(T-T_0-4), & T > T_0+4, \\ \cdots\cdots \\ +\lambda(T-1-1/\lambda), & T > 1+1/\lambda, \\ -\lambda(T-3-1/\lambda), & T > 3+1/\lambda, \end{cases}$$

$$X = \lambda T + \frac{\epsilon\lambda}{2\gamma(1-\lambda^2)}\begin{cases} (1-\lambda)^2T^2, & T > 0, \\ +(1+\lambda)^2(T-2/1+\lambda)^2, & T > 2/1+\lambda, \\ -(1-\lambda)^2(T-2/1-\lambda)^2, & T > 2/1-\lambda, \\ -(1+\lambda)^2(T-4/1+\lambda)^2, & T > 4/1+\lambda, \\ \cdots\cdots \\ -(1-\lambda)^2(T-T_0/1-\lambda)^2, & T > T_0/1-\lambda, \\ -(1+\lambda)^2\left(T-\dfrac{2+T_0}{1+\lambda}\right)^2, & T > \dfrac{2+T_0}{1+\lambda}, \\ \cdots\cdots \end{cases}$$

$$M = \frac{2\lambda\epsilon}{\gamma(1-\lambda^2)} \begin{cases} 0, & T < 1, \\ +(T-1), & T > 1, \\ -(T-3), & T > 3, \\ +(T-5), & T > 5, \\ \cdots\cdots & \\ -(T-1-T_0), & T > 1+T_0, \\ +(T-3-T_0), & T > 3+T_0, \\ \cdots\cdots & \end{cases}$$

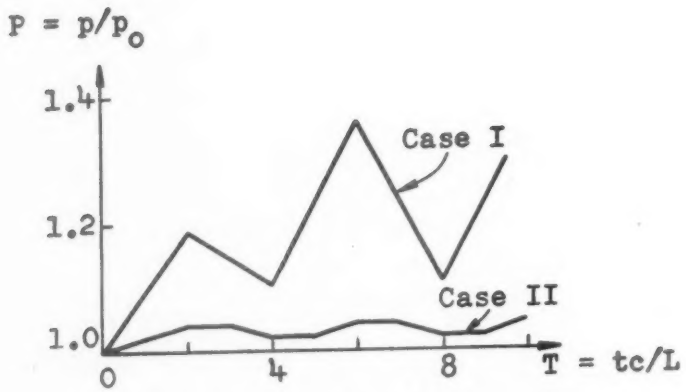

FIG. 3. Pressure at closed end of the tube for cases I and II.

In Figs. 3, 4 and 5 we give graphs for P, X and M for two different sets of possibly typical numerical values for the parameters ϵ, λ and T_0, corresponding respectively to the fast relatively violent phenomena associated with turbulent initial conditions in the tube and to the more mild phenomena associated with static initial conditions in the tube. The two cases are given by, Case I: $\lambda = .3$, $\epsilon = .2$, $T_0 = 6$; Case II: $\lambda = .05$, $\epsilon = .2$, $T_0 = 3$.

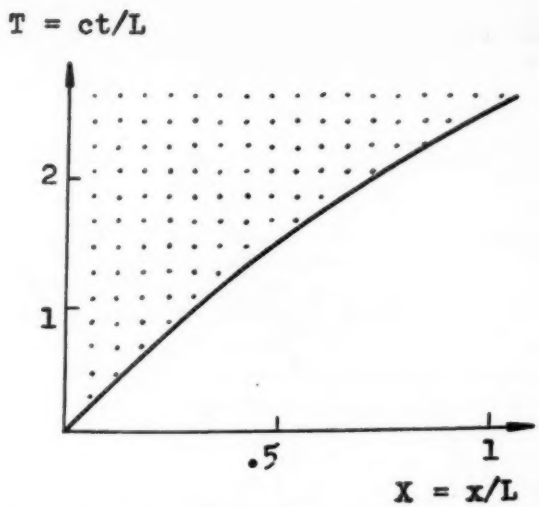

FIG. 4a. Position of the flame front with respect to the tube for Case I.

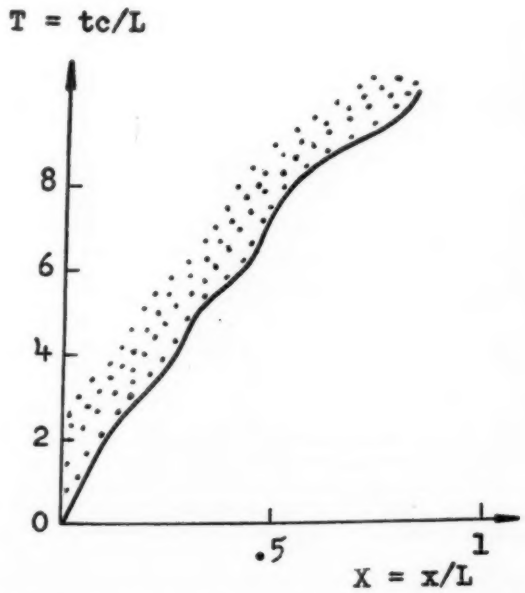

FIG. 4b. Position of the flame front with respect to the tube for Case II.

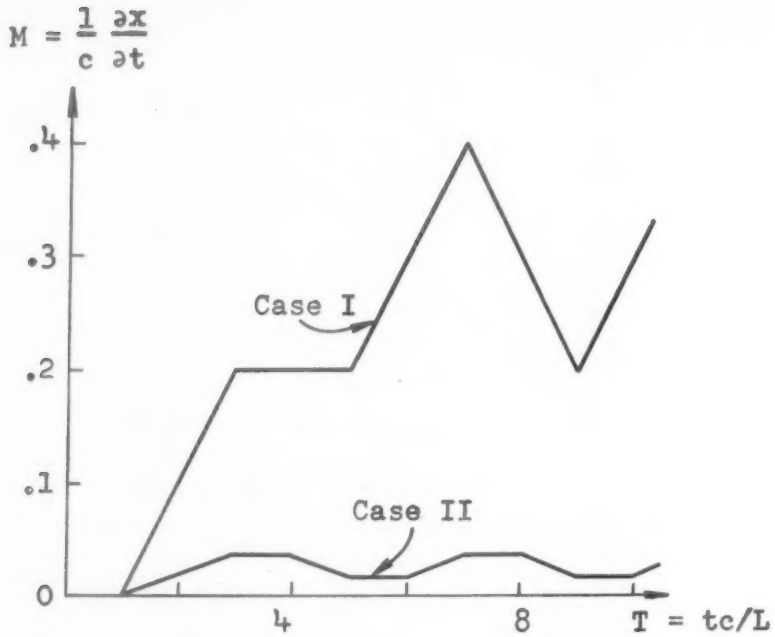

FIG. 5. Mach number $M = \partial x/c\partial t$ of the gas flow at the open end of the tube for Cases I and II.

TORSIONAL RIGIDITY, PRINCIPAL FREQUENCY, ELECTROSTATIC CAPACITY AND SYMMETRIZATION*

BY

GEORGE PÓLYA

Stanford University

Symmetrization is a geometric operation invented by the Swiss mathematician Jacob Steiner. It was observed recently by Mr. Szegö and the author[1] that this geometric operation has definite simple connections with important physical quantities, especially with those mentioned in the title of this paper. This observation leads to many curious consequences some of which may have practical importance.

When the present author was about to present the principal properties of symmetrization in the seminar of Professor Timoshenko, one of the latter's remarks suggested a new version of the method given in the aforementioned paper. This new version will be presented here, as intuitively as possible, and with a numerical illustration, see Sec. 6 (I). As another application, the first proof of an important conjecture of Saint Venant will be given in Sec. 6 (III). For further consequences and several allied results, the reader should refer to the joint papers of Mr. Szegö and the author.[1,2]

1. Definition of symmetrization. Symmetrization applies to plane and solid figures.

Symmetrization in a plane is performed with respect to a specified straight line, called the *line of symmetrization*. For example, symmetrization with respect to the line

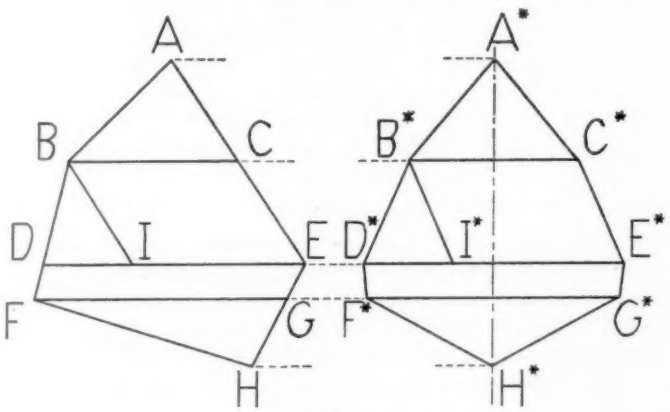

FIG. 1

*Received Dec. 20, 1947. The investigation here presented was carried out within the framework of an Office of Naval Research project directed by Mr. G. Szegö and the author.

[1]G. Pólya and G. Szegö, *Inequalities for the capacity of a condenser*, Amer. J. Math., **67**, 1-32 (1945). See also G. Szegö, *On the capacity of a condenser*, Bulletin Amer. Math. Soc. **51**, 325-350 (1945) and G. Pólya, *Sur la fréquence fondamentale des membranes vibrantes et la résistance élastique des tiges à la torsion*, C. R. Ac. Sci. Paris, **225**, 346-348 (1947).

[2]G. Pólya and G. Szegö, *Approximations and bounds for the electrostatic capacity and for similar physical quantities*. In preparation.

A^*H^* of Fig. 1 transforms the pentagon $ABFHE$ into the octagon $A^*B^*D^*F^*H^*G^*E^*C^*$. The original figure is conceived as consisting of line-segments parallel to each other and perpendicular to the line of symmetrization. Each segment is shifted along its own line into a position where it is bisected by the line of symmetrization. The segments in their new positions form the new, transformed, *symmetrized* figure. Figure 4 shows the effect of two successive symmetrizations. The unsymmetric quadrilateral is symmetrized with respect to the horizontal line I into another quadrilateral (which has, of course, one axis of symmetry). The latter, symmetrized with respect to the vertical line II, is changed into a rhombus. Also, Fig. 3 shows two successive symmetrizations. The rectangle marked with Λ', symmetrized with respect to the line I, is transformed into a rhombus, and this, again, symmetrized with respect to the line II, into another rectangle marked with Λ''.

Symmetrization in space is performed with respect to a specified plane, called "the plane of symmetrization". The original solid is conceived as consisting of parallel line-segments perpendicular to the plane of symmetrization. These segments are shifted, each along its own line, to a position where they are bisected by the plane of symmetrization and form, in their new positions, the transformed solid. For example, in Fig. 2, the plane of symmetrization is perpendicular to the horizontal paper of the drawing which it cuts along A^*C^*. In this Fig. 2, on the left, closed contour lines are

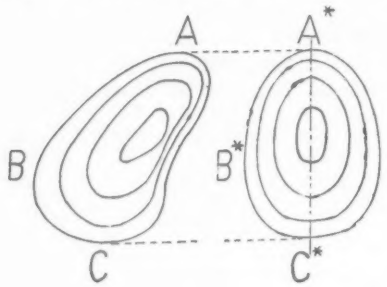

FIG. 2

shown, representing a hill in the terrain. This is transformed by symmetrization with respect to the vertical plane through A^*C^* into a symmetrical hill.[3]

2. Influence of symmetrization on some geometrical quantities. Jacob Steiner found some interesting properties of symmetrization of which we must make essential use later. We summarize these properties in four statements of which the first two are concerned with symmetrization in a plane and the last two with symmetrization in space.

I. *Symmetrization leaves the area unchanged.*

II. *Symmetrization decreases the perimeter.*

III. *Symmetrization leaves the volume unchanged.*

IV. *Symmetrization decreases the surface-area.*

Statements I, II and III are almost immediate. We give a few hints about the elementary geometric considerations which constitute their proof.

[3]In the foregoing explanation we restricted ourselves to figures the boundary of which is not cut in more than two points by any perpendicular to the line, or plane, of symmetrization. It is enough for an introductory treatment to consider this simplest case.

In dealing with symmetrization in a plane, we may restrict our considerations to polygons. In fact, a curved line can be regarded as the limiting case of a polygon. Let us divide the polygon which we intend to symmetrize by parallels passing through its vertices and perpendicular to the line of symmetrization. The polygon will be divided into a few triangles ($\triangle ABC$ and $\triangle FGH$ in Fig. 1) and several trapezoids ($BCED$, $DEGF$). Such a trapezoid is changed into another trapezoid by symmetrization ($BCED$ into $B^*C^*E^*D^*$). These two trapezoids, however, the original and the symmetrized, have the same lower base, the same upper base, the same altitude, and, therefore, the same area. That a triangle is also symmetrized into another with the same area (as $\triangle ABC$ into $\triangle A^*B^*C^*$) is still easier to see. And so assertion I is proved.

In discussing II, we again consider polygons, divided by parallels as before. We must know the elementary proposition that, *of all triangles having the same base and the same area the isosceles has the shortest perimeter.*[4] For example, in Fig. 1, the broken line $B^*A^*C^*$, belonging to an isosceles triangle, is shorter than the corresponding broken line BAC in the non-symmetrized figure. (Consider the line AA^* as a mirror; then $B^*A^*C^*$ is a possible path of light, but not BAC which does not include equal angles with the mirror—and light chooses the shortest path.) We can also compare the sum of the two slanting sides of a trapezoid before and after symmetrization; this sum is shorter after. (For example, $D^*B^* + C^*E^*$, equal to the broken line $D^*B^*I^*$ belonging to the isosceles $\triangle D^*B^*I^*$, is shorter than $DB + CE$, equal to DBI.) Summing up such inequalities over all trapezoids and triangles into which the polygons, before and after symmetrization, are divided, we can justify II.

We do not give a formal proof for III. We just observe that a solid, undergoing symmetrization, can be conceived as consisting of elongated, slender prisms (matches of unequal lengths) all perpendicular to the plane of symmetrization. In the process of symmetrization these prisms (matches) are shifted, each in the line of its altitude, till they are bisected by the plane of symmetrization. Yet such shifting does not alter the volume of the prisms. This remark renders III evident.

The proof of assertion IV by elementary geometry is more difficult. If the reader does not wish to take IV for granted on the basis of the obvious analogy of I and II with III and IV, he may read the following proof which makes use of coordinates and calculus.

We introduce rectangular coordinates. We take the y, z-plane as the plane of symmetrization. (And we take the x, y-plane as the horizontal plane of the drawing. Thus, in Fig. 2, the line A^*C^* represents the y-axis.) A straight line parallel to the x-axis that cuts the surface of the given solid, intersects it at two points which we call (x_1, y, z) and (x_2, y, z). We choose the notation so that $x_1 < x_2$. The area of the surface is expressed by

$$\iint \left[1 + \left(\frac{\partial x_1}{\partial y}\right)^2 + \left(\frac{\partial x_1}{\partial z}\right)^2\right]^{1/2} dy\, dz + \iint \left[1 + \left(\frac{\partial x_2}{\partial y}\right)^2 + \left(\frac{\partial x_2}{\partial z}\right)^2\right]^{1/2} dy\, dz.$$

We introduce the abbreviations

$$\frac{\partial x_1}{\partial y} = Y_1, \qquad \frac{\partial x_1}{\partial z} = Z_1, \qquad \frac{\partial x_2}{\partial y} = Y_2, \qquad \frac{\partial x_2}{\partial z} = Z_2.$$

[4] See e.g. R. Courant and H. Robbins, *What is mathematics?*, pp. 330-333.

Then we can express the area of the surface in the form

$$\iint \{[1 + Y_1^2 + Z_1^2]^{1/2} + [1 + Y_2^2 + Z_2^2]^{1/2}\}\, dy\, dz. \tag{1}$$

After symmetrization, the two points

$$(x_1, y, z) \qquad (x_2, y, z)$$

of the original surface go into the points

$$(-x, y, z) \qquad (x, y, z)$$

of the new, symmetrized surface where

$$x = (x_2 - x_1)/2.$$

The area of the new surface is expressed by

$$2 \iint \left[1 + \left(\frac{\partial x}{\partial y}\right)^2 + \left(\frac{\partial x}{\partial z}\right)^2\right]^{1/2} dy\, dz$$

or, as we find using the abbreviations introduced, by

$$\iint [4 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2]^{1/2}\, dy\, dz. \tag{2}$$

We assert that the integrand in (1) is greater than the integrand in (2). In fact, consider the two vectors with components

$$1, -Y_1, -Z_1 \qquad \text{and} \qquad 1, Y_2, Z_2.$$

Their resultant has the components

$$2, Y_2 - Y_1, Z_2 - Z_1.$$

Yet the sum of the lengths of two vectors can never be less than the length of their resultant. (The sum of two sides of a triangle never less than the third.) Therefore

$$[1 + Y_1^2 + Z_1^2]^{1/2} + [1 + Y_2^2 + Z_2^2]^{1/2} \geq [4 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2]^{1/2}.$$

And so the integrand in the integral (1), expressing the surface-area before symmetrization, is never less than the integrand in the integral (2), expressing the surface-area after symmetrization. This justifies assertion IV.[5]

3. Applications. We wish to draw some consequences from Steiner's theorems recalled in the foregoing section. We use the same coordinate system as before. We take the plane of the drawing in Fig. 2 as our horizontal x, y-plane. That part of this plane which is surrounded by the closed line ABC will be called "the domain $\mathfrak{D}$". We consider a function

$$z = f(x, y) \tag{3}$$

which is positive within the domain $\mathfrak{D}$, but vanishes along the boundary of $\mathfrak{D}$ (the closed line ABC). The surface with equation (3) and the horizontal x, y-plane include a solid, (a "hill" represented by contour lines in Fig. 2) of which $\mathfrak{D}$ is the base. We sym-

[5]The terms "decrease" and "increase" are used here in the wide sense. Strictly speaking, we should have said "never increases" instead of "decreases", etc.

metrize our hill with respect to the y, z-plane which cuts the horizontal plane along the y-axis (the line A^*C^* of Fig. 2). By this symmetrization our hill goes over into a symmetrical hill, of which the base, surrounded by the closed line $A^*B^*C^*$ will be called the domain $\mathfrak{D}^*$. Let

$$z = f^*(x, y) \tag{4}$$

be the equation of the surface that bounds the symmetrized solid (hill) from above. Therefore, $f^*(x, y)$ vanishes along the boundary of $\mathfrak{D}^*$ (the line $A^*B^*C^*$ in Fig. 2 which shows a few more contour lines of the surface (4)). In brief, symmetrization with respect to the y, z-plane transforms the points A, B and C, the line ABC, the domain $\mathfrak{D}$ and the function $f(x, y)$ into the points A^*, B^* and C^*, the line $A^*B^*C^*$, the domain $\mathfrak{D}^*$ and the function $f^*(x, y)$, respectively

Applying the results of the foregoing section to the present case of symmetrization, we observe that the volume remains unchanged (Sec. 2, III), or

$$\iint_{\mathfrak{D}} f \, dx \, dy = \iint_{\mathfrak{D}^*} f^* \, dx \, dy. \tag{5}$$

Yet the area is diminished (Sec. 2, IV), that is

$$\iint_{\mathfrak{D}} [1 + f_x^2 + f_y^2]^{1/2} \, dx \, dy \geq \iint_{\mathfrak{D}^*} [1 + f_x^{*2} + f_y^{*2}]^{1/2} \, dx \, dy \tag{6}$$

where $f_x = \partial f/\partial x$, f_y , f_x^*, and f_y^* denote partial derivatives. Inequality (6) mentions only the curved part of the boundary of the solids compared; but the bases, the area of $\mathfrak{D}$ and that of $\mathfrak{D}^*$, are known to be equal (Sec. 2, I). This remark completes the derivation of (6).

Relation (5) will be useful later. We shall find another useful relation if we apply the inequality (6) not to f, but to ϵf, where ϵ is an arbitrary positive constant. We obtain

$$\iint_{\mathfrak{D}} [1 + \epsilon^2(f_x^2 + f_y^2)]^{1/2} \, dx \, dy \geq \iint_{\mathfrak{D}^*} [1 + \epsilon^2(f_x^{*2} + f_y^{*2})]^{1/2} \, dx \, dy. \tag{7}$$

We suppose that ϵ is sufficiently small, expand both sides of (7) in powers of ϵ, observe that $\mathfrak{D}$ and $\mathfrak{D}^*$ have the same area (Sec. 2, I) and, after obvious cancellation and division by ϵ^2, we obtain

$$\frac{1}{2} \iint_{\mathfrak{D}} (f_x^2 + f_y^2) \, dx \, dy + \cdots \geq \frac{1}{2} \iint_{\mathfrak{D}^*} (f_x^{*2} + f_y^{*2}) \, dx \, dy + \cdots . \tag{8}$$

On both sides, subsequent terms, which are not written, only indicated, contain ϵ^3 or higher powers of ϵ. We let ϵ tend to 0, and obtain finally

$$\iint_{\mathfrak{D}} (f_x^2 + f_y^2) \, dx \, dy \geq \iint_{\mathfrak{D}^*} (f_x^{*2} + f_y^{*2}) \, dx \, dy. \tag{9}$$

This result will turn out to be extremely important in what follows.

By the way, instead of supposing ϵ small, we could have supposed it large. Doing this, we obtain, from (7)

$$\iint_{\mathfrak{D}} [f_x^2 + f_y^2]^{1/2} \, dx \, dy \geq \iint_{\mathfrak{D}^*} [f_x^{*2} + f_y^{*2}]^{1/2} \, dx \, dy.$$

Dividing both sides by the area of $\mathfrak{D}$, or, what is the same, by the area of $\mathfrak{D}^*$, we can state the result by saying that *symmetrization diminishes the average slope*. This observation is quite instructive in showing the smoothing effect of symmetrization, but (5) and (9) will prove more important.

4. Parallel characterization of some physical quantities. Before we can derive the main result, we must make preparations of another kind. The aim of the present section is to characterize certain physical quantities so that their analogy, which is not obvious at first sight, should clearly appear.

I. We consider a homogeneous and isotropic elastic cylinder of which the cross-section is a plane domain $\mathfrak{D}$. When this cylinder is twisted around an axis perpendicular to $\mathfrak{D}$, the couple resisting the twist is $\theta\mu\mathrm{P}$. In this formula θ denotes the twist, that is, the angle of rotation per unit length, μ the modulus of rigidity (or sheer), and P a purely geometric constant, depending on size and shape of the domain $\mathfrak{D}$. I propose to call P the *torsional rigidity of the cross-section* $\mathfrak{D}$. The usual term for $\mu\mathrm{P}$ is torsional rigidity of the elastic cylinder. (There seems to be no generally accepted symbol for torsional rigidity. We use here P, which is the Greek capital corresponding to the initial of "rigidity".)

In investigating the equilibrium position of the twisted cylinder, we can start from the general principle of minimum potential energy. Following the usual reasoning, but using a slightly different notation, we arrive at the following characterization of P, the torsional rigidity of the cross-section $\mathfrak{D}$.

Let $f(x, y)$ be an arbitrary function (just sufficiently "smooth" inside $\mathfrak{D}$) such that

$$f(x, y) = 0 \text{ on the boundary of } \mathfrak{D}. \tag{10}$$

Then

$$\frac{\iint (f_x^2 + f_y^2)\, dx\, dy}{4(\iint f\, dx\, dy)^2} \geq \frac{1}{\mathrm{P}}, \tag{11}$$

where both integrals are extended over the domain $\mathfrak{D}$. Equality in (11) is attained if, and only if, the function $f(x, y)$ is proportional to the stress-function, that is, if $f = \text{const. } U$, where U satisfies the differential equation

$$U_{xx} + U_{yy} + 2 = 0 \tag{12}$$

inside $\mathfrak{D}$ and satisfies the same boundary condition to which f is subjected according to (10). (Of course, $U_{xx} = \partial^2 U/\partial x^2$.) That is, U must vanish along the boundary of $\mathfrak{D}$.

II. We consider a uniform and uniformly stretched elastic membrane, which is fixed along the curve surrounding $\mathfrak{D}$ and coincides with $\mathfrak{D}$ in its equilibrium position. Let Λ denote the frequency of the principal tone of the membrane; the principal tone is the gravest proper tone of the membrane. In computing Λ, we drop a certain physical constant, that depends on the tension and density of the membrane and has the dimension of a velocity. Thus Λ becomes a purely geometric quantity, depending on the shape and size of the domain $\mathfrak{D}$; it is characterized by the inequality

$$\frac{\iint (f_x^2 + f_y^2)\, dx\, dy}{\iint f^2\, dx\, dy} \geq \Lambda^2, \tag{13}$$

where both integrals are extended over the domain $\mathfrak{D}$. This inequality (13) holds for any function f of the same nature as considered before; f is arbitrary inside $\mathfrak{D}$ but

satisfies the boundary condition (10). Equality in (13) is attained if, and only if, the function $f(x, y)$ is proportional to the transverse displacement in the principal vibration, that is, if $f = \text{const. } v$, where v is positive inside $\mathfrak{D}$, satisfies there the differential equation

$$v_{xx} + v_{yy} + \Lambda^2 v = 0 \tag{14}$$

and vanishes along the boundary of $\mathfrak{D}$. Thus, v is also subject to the boundary condition imposed upon f by (10).

III. We consider finally two infinite cylinders, with parallel generatrices, one containing the other. We take the generatrices parallel to the z-axis. Then these cylinders intersect the x, y-plane in two closed curves, one contained in the other. Let us call $\mathfrak{R}$ the ring-shaped plane domain contained between these two curves. (In Fig. 2, we take the curve ABC as outer boundary of $\mathfrak{R}$, and the innermost of the closed lines drawn inside ABC as inner boundary.) We conceive the two cylinders as conductors of electricity, both having constant potential; the potential of the outer cylinder is 0, that of the inner one different from 0. The cylinders form a condenser and we call C the capacity of this condenser per unit height.

Also C can be characterized by an inequality. Let $f(x, y)$ be an arbitrary function (just smooth inside $\mathfrak{R}$) and such that

$f(x, y) = 0$ along the outer,

$f(x, y) = f_1$ along the inner

boundary of $\mathfrak{R}$ where f_1 is a constant, different from 0. Then

$$\frac{\iint (f_x^2 + f_y^2)\, dx\, dy}{4\pi f_1^2} \geq C; \tag{15}$$

the integral is extended over the ring-shaped domain $\mathfrak{R}$. Inequality (15) holds for an arbitrary function $f(x, y)$ of the described type. Equality in (15) is attained if, and only if, $f(x, y)$ is proportional to the electrostatic potential, that is, if $f = \text{const. } w$, where w satisfies the differential equation

$$w_{xx} + w_{yy} = 0 \tag{16}$$

inside $\mathfrak{R}$ and boundary conditions of the same nature as $f(x, y)$:

$$w = 0 \qquad \text{on outer,}$$

$$w = 1 \qquad \text{on inner,}$$

boundary of $\mathfrak{R}$.

5. Influence of symmetrization on some physical quantities. After the preparations in the foregoing sections, we are now in a position to derive the following statements which are analogous to Steiner's Theorems discussed in Sec. 2.

I. *Symmetrization increases the torsional rigidity.*

II. *Symmetrization decreases the principal frequency.*

III. *Symmetrization decreases the electrostatic capacity.*

In brief, we assert that symmetrization decreases all three quantities, $1/P$, Λ and C, characterized by the parallel inequalities (11), (13) and (15) and connected with the corresponding differential equations (12), (14) and (16). The analogy so disclosed is an essential part of the proof. In fact, after having considered carefully the derivation of Theorem I, concerned with torsional rigidity, we shall need only a few additional remarks about II and III.

We consider now two elastic shafts of which the cross-sections are displayed in Fig. 2: the domain $\mathfrak{D}$, bounded by the curve ABC, and $\mathfrak{D}^*$, bounded by $A^*B^*C^*$. Let P and P* denote the torsional rigidities corresponding to $\mathfrak{D}$ and $\mathfrak{D}^*$, respectively.

As we have discussed in Sec. 4 (I), there is a function $f(x, y)$, satisfying the boundary condition (10), and such that

$$\frac{1}{\mathrm{P}} = \frac{\iint_{\mathfrak{D}} (f_x^2 + f_y^2)\, dx\, dy}{4(\iint_{\mathfrak{D}} f\, dx\, dy)^2}. \tag{17}$$

This f is the true stress-function corresponding to the cross-section $\mathfrak{D}$, or proportional to it. Thus, in Fig. 2, the contour lines inside ABC represent the actual stress distribution in the twisted shaft with cross-section $\mathfrak{D}$.

Now, we symmetrize the domain $\mathfrak{D}$ and the function $f(x, y)$ as described in Sec. 3, and obtain $\mathfrak{D}^*$ and $f^*(x, y)$, respectively. The symmetrized domain $\mathfrak{D}^*$ is included by the line $A^*B^*C^*$ in Fig. 2. The symmetrized function $f^*(x, y)$ is positive inside $\mathfrak{D}^*$ and vanishes along its boundary. From the relations (5) and (9), derived in Sec. 3, we conclude that

$$\frac{\iint_{\mathfrak{D}} (f_x^2 + f_y^2)\, dx\, dy}{4(\iint_{\mathfrak{D}} f\, dx\, dy)^2} \geq \frac{\iint_{\mathfrak{D}^*} (f_x^{*2} + f_y^{*2})\, dx\, dy}{4(\iint_{\mathfrak{D}^*} f^*\, dx\, dy)^2} \tag{18}$$

The function $f(x, y)$ was the true stress-function for $\mathfrak{D}$ but, of course, we do not know whether f^* is or is not the stress function for $\mathfrak{D}$.* (Very likely it is not. We have no reason to suppose that the contour lines inside $A^*B^*C^*$, obtained by symmetrization, represent the true stress distribution.) Yet, at any rate, we can apply the inequality (11) characterizing the torsional rigidity to the case of the domain $\mathfrak{D}^*$, and so we obtain that

$$\frac{\iint_{\mathfrak{D}^*} (f_x^{*2} + f_y^{*2})\, dx\, dy}{4(\iint_{\mathfrak{D}^*} f^*\, dx\, dy)^2} \geq \frac{1}{\mathrm{P}^*}. \tag{19}$$

Joining (17), (18) and (19), we conclude that

$$\frac{1}{\mathrm{P}} \geq \frac{1}{\mathrm{P}^*}$$

or

$$\mathrm{P}^* \geq \mathrm{P}.$$

We have justified our assertion I.

Only small modifications of the foregoing reasoning are needed to handle the cases of principal frequency and electrostatic capacity, and to derive the corresponding Theorems II and III. To prove II, we must use, instead of (5), that

$$\iint_{\mathfrak{D}} f^2\, dx\, dy = \iint_{\mathfrak{D}^*} f^{*2}\, dx\, dy. \tag{20}$$

Yet this is evident. Indeed, when the surface with equation

$$z = f(x, y)$$

is symmetrized, the surface with equation

$$z = [f(x, y)]^2$$

is symmetrized along with it—a fact that we could express with the equation

$$(f^2)^* = (f^*)^2.$$

Therefore, (20) follows from the proposition that the volume is preserved by symmetrization (Sec. 2, III).

6. Applications. A few examples shall illustrate the general results of the foregoing section.

I. For a rectangle, both the torsional rigidity and the principal frequency have been found, but for a rhombus none of these two important quantities has been exactly expressed yet as far as I know. From Fig. 3, however, we can derive limits for them.

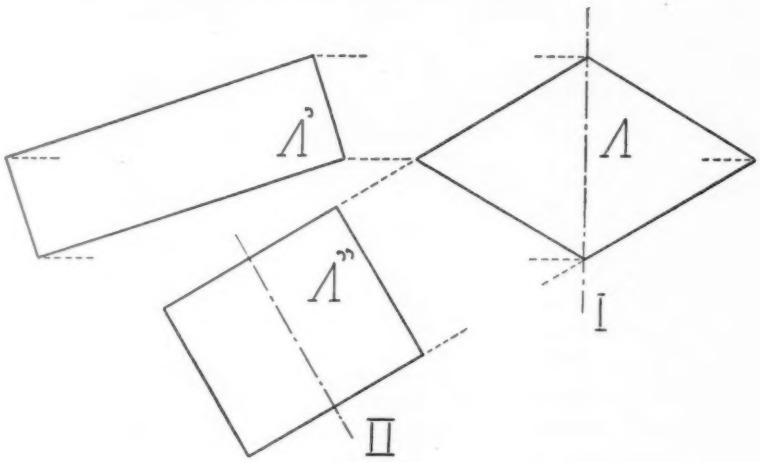

FIG. 3

Let Λ, Λ' and Λ'' denote the principal frequencies of the figures marked with these letters, respectively. Symmetrized with respect to the straight line I, the rectangle with principal frequency Λ' is transformed into a rhombus, and the latter, symmetrized with respect to II, into the other rectangle. Symmetrization, however, decreases the fundamental frequency (Sec. 5, II). Therefore

$$\Lambda' > \Lambda > \Lambda'' \tag{21}$$

Let d denote the shorter diagonal and δ the acute angle of the rhombus. A little elementary geometry is enough to compute the sides of the two rectangles, after which (21) yields the following bounds for Λ, the principal frequency of the rhombus:

$$\frac{2\pi}{d} > \Lambda > \frac{\pi}{d}\frac{(1+\sin^2\delta)^{1/2}}{\cos(\delta/2)}. \tag{22}$$

The lower bound is much nearer to the truth. In fact, it attains the exact value for Λ in both extreme cases, for $\delta = \pi/2$ (square) and for $\delta = 0$ (infinite strip). The upper bound yields the exact value only for $\delta = \pi/2$ (square).

[By other means, namely by conformal representation, as it will be explained in

the joint paper quoted under[2], we can derive an upper bound for Λ which is much better than the bound $2\pi/d$, except in cases in which δ is very near to $\pi/2$ and the rhombus is almost a square:

$$\Lambda < \frac{\pi}{d}\frac{j}{2\pi^{1/2}}\,\Gamma\left(\frac{1}{2}-\frac{\delta}{2\pi}\right)\Big/\,\Gamma\left(1-\frac{\delta}{2\pi}\right); \tag{23}$$

$j = 2.4048$ is the first positive zero of the Bessel function $J_0(x)$.

The following short table shows which degree of approximation we can obtain by combining the upper bound (23) with the lower bound of (22).

δ	$\dfrac{(1+\sin^2\delta)^{1/2}}{\cos(\delta/2)}$	$\dfrac{j\Gamma[(\pi-\delta)/2\pi]}{2\pi^{1/2}\Gamma[(2\pi-\delta)/2\pi]}$
90°	2.0000	2.0071
89	1.9826	1.9898
88	1.9654	1.9728
87	1.9483	1.9562
86	1.9313	1.9400
85	1.9145	1.9241
80	1.8322	1.8491
75	1.7525	1.7810
70	1.6752	1.7190
65	1.6002	1.6621
60	1.5273	1.6100
45	1.3256	1.4758
30	1.1575	1.3674
15	1.0419	1.2778
0	1.0000	1.2024

The first column expresses δ in degrees, the third in radians.

Taking the mean between the two bounds, we are certain to obtain the value of Λ for a rhombus with an error of less than 1% as long as δ is not less than 75°.]

Also the torsional rigidity of a rhombic cross-section can be included between two limits by using Fig. 3.

II. Figure 4 shows how a quadrilateral can be transformed into a rhombus by two successive symmetrizations. Now, by Fig. 3, the rhombus so obtained can be symmetrized into a rectangle, the rectangle into another rhombus, this into another rectangle, and so on. As Fig. 3 suggests, the rectangles successively generated by this process become more and more square-like, and it is easy to prove that we obtain a square in the limit. Now, each of these successive symmetrizations preserves the area, increases the torsional rigidity and decreases the principal frequency. Therefore, our process shows: *Of all quadrilaterals with a given area, the square has the maximum torsional rigidity and the minimum principal frequency.*

A similar, but more obvious reasoning applies to triangles; compare, in Fig. 1, the general $\triangle ABC$ with the isosceles $\triangle A^*B^*C^*$. An appropriate process of successive symmetrization shows that *of all triangles with a given area, the equilateral triangle has the highest torsional rigidity and the lowest principal frequency.* Yet we could not handle in

this manner the case of pentagons. Symmetrizing an irregular pentagon, we cannot avoid changing it into some polygon with more than five sides. (In Fig. 1, a pentagon is changed into an octagon.)

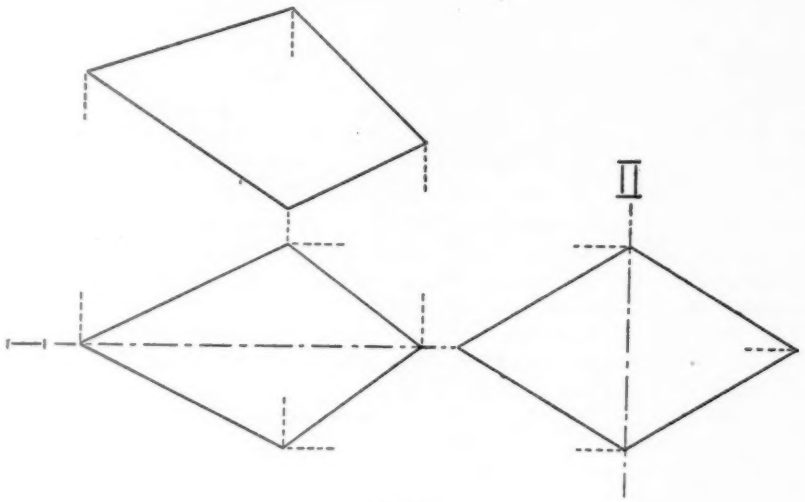

FIG. 4

It is natural to suspect that the propositions just proved about the equilateral triangle and the square can be extended to regular polygons with more than four sides. Yet the difficulty that we face when attempting the extension to pentagons can make us realize that our theorems on symmetrization are far from being trivial.

III. A circle is symmetric with respect to any of its diameters and remains unchanged by symmetrization with respect to any line whatsoever. On the other hand, by an appropriate (infinite) sequence of symmetrizations, any plane figure can be changed into a circle ultimately. (This is plausible, and has been proved.) Therefore, *of all plane domains with a given area, the circle has the maximum torsional rigidity and the minimum principal frequency.*

The second part of this statement, concerning principal frequency, has been conjectured by Lord Rayleigh, and later proved by Faber. The first part, concerning torsional rigidity, has been clearly foreseen by Saint-Venant who supported it by convincing physical considerations which, however, do not amount to a mathematical proof, and, to my knowledge, the first such proof is the one just presented.

TWO PROBLEMS OF MIXED TYPE FOR THE DAMPED WAVE EQUATION*

BY

D. G. BOURGIN

University of Illinois

This paper is primarily concerned with a mixed problem of somewhat unconventional form for the one dimensional damped wave equation. Although the treatment given is rigorous, the main interest of this work lies in the explicit determination of the solution in closed form and in the methods used. The method of attack pivots on the use of images (or co-images) associated with a hyperbolic metric. The utility of such extensions of the method of images, so useful for equations of elliptic type, was first pointed out by Volterra.[1] However, the literature contains scant references to the application[2] and development of this idea. For the purpose of cataloguing the nature of our addition to the class of explicitly soluble problems it may be pointed out that the fundamental integral equations for the main problem are not of convolution type (*cf.* Sec. 7).

The main problem may be given significant physical interpretation in various fields. For instance, it covers the vibration of a damped string whose length varies linearly with time. One end moves in a vertical slot and the other is coupled to a driving mechanism through an inertialess damped spring. Alternatively we may consider one dimensional gas motion (or sound propagation) in a cylinder of large radius. One end is closed. The other moves at constant rate and is a friction damped elastic diaphragm driven by an exterior mechanism. We may think of our problem as that of a telegrapher's equation problem for an electric cable of steadily increasing length. A fourth application is to problems connected with Schuster's theory of the radiation field in a moving radiating atmosphere in the formulation given by Chandrasekhar. Indeed the writer's interest in this investigation was aroused by a manuscript which Dr. Chandrasekhar communicated to him.[3] The second problem treated in this article is already a generalization of that of Chandrasekhar's, though the developments are simpler than those for our main problem.

1. Preliminaries. We shall be interested in a boundary value problem associated with

$$z_{tt} - z_{xx} + z = L(z) = 0. \tag{1.0}$$

Before we state it, it seems desirable to indicate in brief the basic theory that we draw on.

Let z and u be solutions of (1.0), then for a continuous rectifiable simple closed curve, Γ, bounding a domain S we have

*Received Sept. 22, 1947.

[1] V. Volterra, *Leçons sur l'intégration des équations différentielles aux dérivées partielles*, A. Herman et Fils; Paris, 1912.

[2] In the thesis just completed by Mr. A. Charnes (University of Illinois) this method is of central importance in analyzing the influence of a polygonal section type of fuselage on the wing potential for the linearized supersonic flow.

[3] His interesting method has since been published: S. Chandrasekhar, *A new type of boundary value problem in hyperbolic equations*, Proc. Cambridge Philos. Soc. **42**, 250-260 (1946). Actually, the generalization of Chandrasekhar's boundary conditions to problem 2 would give somewhat different conditions than those used in this paper, but these are just as easily handled. On this point *cf.* footnote 4.

$$0 = \iint_S [uL(z) - zL(u)]\, dS = \int_\Gamma [u(z_x t_s + z_t x_s) - z(u_x t_s + u_t x_s)]\, ds, \quad (1.01)$$

for a *clockwise sense* of description of Γ. Besides the Euclidean metric used for ds we associate a hyperbolic metric with (1.01), namely

$$d\sigma^2 = dt^2 - dx^2. \quad (1.02)$$

The characteristics of (1.0) are the null lines for $d\sigma$, namely $t + x = C$, $t - x = C$. We shall refer to the members of these two families as plus characteristics and minus characteristics respectively. A pair of vectors orthogonal with respect to this metric are said to be co-normal. Thus the vector δx, δt is co-normal to the tangent vector (to Γ) dx, dt provided

$$\delta t\, dt - \delta x\, dx = 0.$$

We denote the co-normal direction by ν. That is to say, we write

$$\frac{\delta x}{\delta \nu} = \frac{dt}{ds}, \qquad \frac{\delta t}{\delta \nu} = \frac{dx}{ds}$$

If Γ is given by $f(x, t) = 0$ then of course

$$\begin{aligned} \frac{dx}{ds} &= \pm f_t/(f_x^2 + f_t^2)^{1/2} = \frac{\partial t}{\partial \nu}, \\ \frac{dt}{ds} &= \mp f_x/(f_x^2 + f_t^2)^{1/2} = \frac{\partial x}{\partial \nu}. \end{aligned} \quad (1.03)$$

The sign choice is determined by the order on the curve. Thus (1.01) becomes

$$\int_\Gamma \left(v \frac{\partial z}{\partial \nu} - z \frac{\partial v}{\partial \nu} \right) ds = 0 \quad (1.04)$$

We shall take for v the Riemann function

$$u(x, t; x_1, t_1) = J_0((t - t_1)^2 - (x - x_1)^2)^{1/2}. \quad (1.05)$$

For our purpose the significant properties of u are first that in the sector expanding downwards bounded by $t \pm x = t_1 \pm x_1$, it satisfies (1.0) and second that u has the constant value 1 on the two sides of this sector. It is then clear that the tangential derivative of u along these sides vanishes. On the plus characteristic $t + x = C$ we have from (1.03) that

$$\left(\frac{dx}{ds}, \frac{dt}{ds} \right) = (1, -1) = \left(-\frac{\partial x}{\partial \nu}, \frac{\partial t}{\partial \nu} \right).$$

That is to say, for these characteristics

$$\frac{\partial}{\partial \nu} = -\frac{d}{ds}.$$

On the other hand (1.03) shows that for the minus characteristic $t - x = C$ we get

$$\frac{\partial}{\partial \nu} = \frac{d}{ds}.$$

Hence if Γ includes a segment of a characteristic we have as the contribution to (1.04), with $s_2 > s_1$,

$$\int_{s_1}^{s_2} \frac{\partial z}{\partial \nu}\, ds = \pm z \Big|_{s_1}^{s_2}, \tag{1.06}$$

according as the characteristic is a minus or a plus one.

Our concern is with the solution of (1.0) restricted by boundary conditions on $t = 0$, $P : t + \lambda x = 0$ and a line P' which is $x = l$ for our main problem and $t + \lambda x = \lambda l$ for our second problem. In view of the linearity of our system we can express the general solution as a sum of three solutions each determined by 0 data on two of the boundaries. Thus consider the solution of (1.0) subject to[4]

$$\begin{aligned} &(A) : z_\nu = \gamma(x, t) \text{ on } P', \\ &(B) : z = A(x),\ z_t = B(x) \text{ on } t = 0, \\ &(C) : z_\nu = hz(x, t) + \psi(x, t) \text{ on } P. \end{aligned}$$

Let z^A be the solution of (1.0) subject to vanishing Cauchy data on $t = 0$, $z_\nu = 0$ on P and $z_\nu = \gamma(x, t)$ on P'. Let z^B be the solution of (1.0) subject to $z = A(x)$, $z_t = B(x)$ on $t = 0$; $z_\nu = 0$ on P' and $z_\nu = 0$ on P. Let z^C be the solution of (1.0) under the conditions $z_\nu = 0$ on P', vanishing Cauchy conditions on $t = 0$ and $z_\nu = hz(x, t) + \psi(x, t) + h(z^A + z^B)$ on P. It is easily verified that the solution of (1.0) restricted by (A), (B) and (C) is then given by $z = z^A + z^B + z^C$.

It is clear that the crux of the general problem is the determination of z^C for z^A and z^B can be determined by a simpler form of the methods of this paper (or otherwise). Thus we shall concern ourselves only with the solution of (1.0) subject to

$$\begin{aligned} &z_\nu = 0 \text{ on } P', \\ &z = z_t = 0 \text{ on } t = 0, \\ &z_\nu = hz(x, t) + \psi(x, t) \text{ on } P. \end{aligned} \tag{1.07}$$

We make the further assumptions that $\psi(0, 0) = 0$ and that on P, $d/ds\ \psi(0, 0) = 0$ and[5] $\psi(x, t) \varepsilon C'$. We suppose also that $\lambda > 1$ and $h \geq 0$ though this last restriction is not significant for our analysis.

2. Geometric considerations. On P consider s increasing with t. This will be the situation in our developments. Then dx/ds is negative and dt/ds is positive. From (1.03) we get

$$\begin{aligned} \frac{\partial x}{\partial \nu} &= \lambda/(\lambda^2 + 1)^{1/2} \\ \frac{\partial t}{\partial \nu} &= -(\lambda^2 + 1)^{-1/2} \end{aligned} \tag{2.0}$$

[4] Chandrasekhar's problem corresponds to the case $\lambda = \infty$. His conditions would amount to changing hz to z_s in (C) and z_ν to z in (A). The last modification is evidently trivial for the corresponding z^A is just as easily found. The first change has the effect of replacing h by p in all the transforms arising in Problem 2. Since, however, the inverse transforms for this case are direct and essentially on record, Cf. 8, the problem is actually simpler than before.

[5] C^n denotes the class of functions continuous up to and including the nth derivatives with respect to the independent variables (here a single one).

The key of our method is the use of images. The reader will find a figure helpful. Consider the image in P of the point (x_0 , t_0) with respect to the hyperbolic metric (1.02). Denote this image by L. Then the hyperbolic distances from any point of P to L and to (x_0 , t_0) are equal. Thus if L is (x', t') then for all x,

$$(\lambda x + t_0)^2 - (x - x_0)^2 = (\lambda x + t')^2 - (x - x')^2, \tag{2.01}$$

whence

$$x' = -(2\lambda t_0 + (\lambda^2 + 1)x_0)/(\lambda^2 + 1),$$

$$t' = \quad (2\lambda x_0 + (\lambda^2 + 1)x_0)/(\lambda^2 - 1).$$

It is convenient to use the representation

$$\cosh \alpha = (\lambda^2 + 1)/(\lambda^2 - 1),$$

$$\sinh \alpha = 2\lambda/\lambda^2 - 1. \tag{2.02}$$

Thus

$$x' = -(x_0 \cosh \alpha + t_0 \sinh \alpha),$$

$$t' = \quad x_0 \sinh \alpha + t_0 \cosh \alpha. \tag{2.03}$$

We denote the image of x_0 , t_0 in P' by R. The calculations of this section refer to the main problem. Thus

$$R : (2l - x_0 , t_0).$$

The image of L in P' is denoted by LR. That of R in P by RL, etc. Hence the point RL has coordinates

$$x' = 2l + x_0 \cosh \alpha + t_0 \sinh \alpha,$$

$$t' = \quad x_0 \sinh \alpha + t_0 \cosh \alpha. \tag{2.04}$$

It is plain that $\cdots R$ and the image corresponding to this symbol with an added L on the right, namely $\cdots RL$, have equal hyperbolic distances from any point on P. A similar remark holds for $\cdots L$ and $\cdots LR$ with respect to P'. Accordingly the function in (1.05) with x_1 , t_1 corresponding to $\cdots R$ and that with x_1 , t_1 corresponding to $\cdots RL$ have equal values on P. Moreover, the derivatives with respect to ν of these two u's are equal, but of opposite sign on P. Thus with $(t - t_0)^2 - (x - x_0)^2 = (t - t_1)^2 - (x - x_1)^2 = \mu^2$ for x, t on P we have

$$(\lambda^2 + 1)^{1/2} \frac{\partial}{\partial \nu} J_0[(t - t_0)^2 - (x - x_0)^2]^{1/2} = \frac{1}{\mu} \frac{d}{d\mu} (J_0(\mu)) (\lambda x_0 + t_0)$$

We have therefore merely to show

$$\lambda x_0 + t_0 = -(\lambda x_1 + t_1)$$

for $L : x_1 , t_1$. However, this is an immediate consequence of (2.03).

Denote by J the combination LR and by J^n the nth iterate. Denote by K^{n+1} the point KJ^n, where $K = RJ^0 = R$. Let $\mathbf{M}$ be the matrix

$$\begin{pmatrix} \cosh \alpha & \sinh \alpha \\ \sinh \alpha & \cosh \alpha \end{pmatrix}. \tag{2.05}$$

We may give (a, b) the dual interpretation of a point (or vector) and a one rowed matrix. Thus with the familiar algorithm for matrix multiplication, we may write (2.04) as

$$J : (x_1 t_1) = (x_0 t_0)\mathbf{M} + (2l0).$$

Suppose first $l = 0$. Then the coordinates of J^n are given by

$$(x_n t_n) = (x_0 t_0)\mathbf{M}^n.$$

Since $\mathbf{M}$ induces a rotation of angle α in the sense of the hyperbolic metric, i.e. a Lorentz rotation, it follows at once that

$$\mathbf{M}^n = \begin{pmatrix} \cosh n\alpha & \sinh n\alpha \\ \sinh n\alpha & \cosh n\alpha \end{pmatrix} \tag{2.06}$$

A direct proof of (2.06) proceeds by induction, using the addition laws for the hyperbolic functions.

The contribution of the l terms may be obtained by assuming $x_0 = t_0 = 0$ and $l \neq 0$. Let l_n , l'_n be the l terms in the x and t coordinates of J^n. Then according to (2.03) and (2.04)

$$l_{n+1} = 2l + l_n \cosh \alpha$$

$$l'_{n+1} = l'_n \sinh \alpha$$

That is to say

$$(l_{n+1} l'_{n+1}) = (2l0) + (l_n l'_n)\mathbf{M}.$$

Hence

$$(l_n l'_n) = (2l0) \sum_0^{n-1} \mathbf{M}^j = (2l0) \begin{pmatrix} e_n 0 \\ 0 0_n \end{pmatrix} \tag{2.06}$$

where e_n and 0_n are the even and odd parts of $\sum_0^{n-1} \mathbf{M}^j$ and are equal to $\sum_0^{n-1} \cosh j\alpha$ and $\sum_0^{n-1} \sinh j\alpha$ respectively. Hence $J^n : X_n , T_n$ is given by

$$\begin{aligned} X_n &= e_n l + x_0 \cosh n\alpha + t_0 \sinh n\alpha, \\ T_n &= 0_n l + x_0 \sinh n\alpha + t_0 \cosh n\alpha. \end{aligned} \tag{2.07}$$

The coordinates of K^n are designated by X'_n , T'_n . It will be clear that K^n arises from J^{n-1} by the substitution of $2l - x_0$ for x_0 . We observe that $\cosh j\alpha$ and $\sinh j\alpha$ are even and odd functions respectively as regards both α and λ. Accordingly the reader may verify the following formulae convenient for our purpose

$$\begin{aligned} e_n &= \lambda \sinh n\alpha + (1 - \cosh n\alpha), \\ 0_n &= \lambda (\cosh n\alpha - 1) - \sinh n\alpha. \end{aligned} \tag{2.08}$$

It is convenient to change from the coordinate system (x, t) to another, (w, v), in which the coordinate lines are again conormal, i.e. orthogonal in the metric of (1.02).

We require that $v = 0$ give P and that the transformation from (x, t) to (w, v) preserve the hyperbolic distance. It is then easy to see that we must have

$$\begin{aligned} w &= \rho(\lambda t + x), \\ v &= \rho(\lambda x + t). \end{aligned} \tag{2.09}$$

We require further that w retain its previous significance as Euclidean distance on P. Hence for x, t on P we have

$$\begin{aligned} t &= \lambda w/(\lambda^2 + 1)^{1/2} \\ &= \lambda w/(\rho(\lambda^2 - 1)). \end{aligned}$$

Accordingly $\rho = (\lambda^2 + 1)^{1/2}/(\lambda^2 - 1)$ and

$$d\sigma^2 = \operatorname{sech} \alpha(dw^2 - dv^2). \tag{2.10}$$

Hereafter β^2 is written for sech α with $\beta > 0$.

The w, v coordinates of x_0 , t_0 are designated by w_0 , v_0 . Those of X_n , T_n and X_n' , T_n' by W_n , V_n and W_n' , V_n' respectively. When $v_0 = 0$ the distinction between primed and unprimed letters fails and we denote them by w_n and v_n in the obvious correspondence. We find in a straightforward way that

$$\begin{aligned} W_n &= (\cosh n\alpha - 1)l(\lambda^2 + 1)^{1/2} + w_0 \cosh n\alpha + v_0 \sinh n\alpha. \\ V_n &= \sinh n\alpha l\,(\lambda^2 + 1)^{1/2} + v_0 \cosh n\alpha + w_0 \sinh n\alpha. \\ W_n' &= (\cosh n\alpha - 1)l(\lambda^2 + 1)^{1/2} + w_0 \cosh n\alpha - v_0 \sinh n\alpha. \\ V_n' &= \sinh n\alpha l\,(\lambda^2 + 1)^{1/2} + w_0 \sinh n\alpha - v_0 \cosh n\alpha. \end{aligned} \tag{2.11}$$

The minus characteristic through W_n , V_n cuts $v = 0$ at $w = w^n$ and $x = l$ at $t = t^n$. We define w'^n, v'^n similarly with respect to the minus characteristic through W_n' , V_n' . Observe that J^{n-1} has $J^{n-1}L$ as its image in P and $J^{n-1}L$ has J^n as its image in P'. It therefore follows from the definition of images that the plus characteristic through $J^{n-1}L$ cuts P at $w = w^{n-1}$ and $x = l$ at $t = t^n$. Indeed the hyperbolic distance of $(w^{n-1}, 0)$ from either J^{n-1} or $J^{n-1}L$ is 0 since characteristics are null geodesics. Analogous remarks hold for the left image points of $K^n : W_n'$, V_n' .

It is manifest from (2.09) that in the w, v coordinates the characteristics are $w \pm v = C$ and that plus characteristics in the x, t system are again plus in the w, v system. Accordingly[6]

$$\begin{aligned} w^0 &= w_0 - v_0 , \\ w^n &= W_n - V_n , \\ w'^n &= W_n' - V_n' . \end{aligned} \tag{2.12}$$

[6]Repeated tacit use is made of (2.12), particularly as regards changing from the sub to superscript 0 and vice versa.

For x, t on P we have[7]

$$J_0[(t - T_n)^2 - (x - X_n)^2]^{1/2} = J_0\beta[(w - W_n)^2 - (V_n)^2]^{1/2} \tag{2.13}$$

and a corresponding expression in W_n', V_n' if the image point is X_n', T_n'.

3. The associated integral equations. Let γ designate the curve P, P' (for $t \geq 0$) and P'' where P'' is the segment of $t = 0$ connecting P and P'. Consider w_0, v_0 or any of the image points L, R, J^n, K^n, J^nL, K^nL as an *apex* for the plus and minus characteristic rays. We need consider only the case where one, at least, of these rays intersects γ. Then a *convex polygon* is uniquely determined by these rays and γ. The boundary of this polygon is denoted by Γ. For convenience we shall use an obvious mixed notation. Thus if the first argument of $z(\ ,\ )$ is x, the other is t, if the first is w the second is v. Suppose w_0, v_0 is the apex. The associated polygon may be five sided. Assume this is indeed the case for the w_0, v_0 chosen. Then by reason of (1.07) and the properties of the Riemann function, (1.04) becomes

$$\begin{aligned} &2z(w_0, v_0) - z(w^0, 0) - z(l, t') \\ &\quad + \int_0^{w^0} \left[J_0\beta(w^0 - w)z_v(w, 0) - z(w, 0)\frac{\partial}{\partial v} J_0\beta(w^0 - w) \right|_P \Big] dw \qquad (3.00) \\ &\quad + \int_0^{t'} z(l, t)\frac{\partial}{\partial x} J_0[(t - t_0)^2 - (l - x_0)^2]^{1/2}\, dt = 0. \end{aligned}$$

If an image point is taken as apex then the associated polygon is at most a quadrilateral. Assume this optimum case. Then the vertices of the polygon are at $(\omega_n, 0)$, (l, τ_n), $(0, 0)$ and $(l, 0)$ where the last refers to the (x, t) coordinates. Here ω_n, τ_n is either w^n, t^n; w'^n, t'^n; w^n, t^{n+1} or w'^n, t'^{n+1} depending on whether J^n, K^n, J^nL or K^nL is the image point selected as apex.

$$\begin{aligned} &z(\omega_n, 0) - z(l, \tau_n) + \int_0^{\omega_n} \left[J_0(\beta D)\frac{\partial}{\partial v} z(w, 0) - z(w, 0)\frac{\partial}{\partial v} J_0(\beta D) \right|_P \Big] dw \\ &\quad + \int_0^{\tau_n} z(l, t)\frac{\partial}{\partial x} J_0(\beta D) \Big|_{x=l} dt = 0, \qquad n \geq 1. \end{aligned} \tag{3.01}$$

The argument βD of $J_0(\beta D)$ is β times the hyperbolic distance from the apex point to P or P' according as the first or second integral is considered in (3.01). We may now waive the requirement that ω_n, τ_n be non negative in (3.0) or (3.01) provided it be understood that a *negative integration limit means the integral is* 0 and that $z(w, 0)$ and $z(l, t)$ are 0 for w or t negative.

We sum on n, the left hand members of (3.01) plus (3.0). This is a finite sum since starting with arbitrary w_0, v_0, eventually ω_n and τ_n are negative. Hence if necessary the order of summing may be modified at will. Manifestly the integrated terms balance out in pairs and one is left with $2z(w_0, v_0)$. The integrals enter the sum in pairs having the same upper limit. Specifically if the upper limit is w^n then the pair of integrals arises

[7] For typographical reasons we generally omit a parenthesis and write $J_0\beta[(w - a)^2 - b^2]^{1/2}$ in place of $J_0\{\beta[(w - a)^2 - b^2]^{1/2}\}$.

from the pair of images J^n, J^nL. If the upper limit is w'^n then the image pair is K^n, K^nL, etc. With respect to t^n the pairs are J^nL, J^{n+1}. Now the Riemann functions for the pair J^n, J^nL for instance have equal values on P and opposite values for the co-normal derivatives on P. Similar remarks hold for P' or for the other image pairs. Accordingly the terms involving the integrands balance out and all integrals along P' balance out. Our final result is, then,

$$-z(w_0 , v_0) = \int_0^{w^0} J_0\beta[(w - W_0)^2 - (V_0)^2]^{1/2} \left.\frac{\partial z}{\partial v}\right|_P dw$$

$$+ \sum_{n=1} \int_0^{w^n} J_0\beta[(w - W_n)^2 - (V_n)^2]^{1/2} \left.\frac{\partial z}{\partial v}\right|_P dw \tag{3.02}$$

$$+ \sum_{n=1} \int_0^{w'^n} J_0\beta[(w - W'_n)^2 - (V'_n)^2]^{1/2} \left.\frac{\partial z}{\partial v}\right|_P dw.$$

It is convenient for our method of attack on (3.02) that one limit of the integrals occurring be the same for all integrals. Therefore replace w by $w^0 - w$ in the integral equation above and use (1.07). This leads to our basic formulation

$$-z(w_0 , v_0) = \sum_{n=0} \int_{w^n}^{w^0} J_0\beta[(w - w^0 + W_n)^2 - (V_n)^2]^{1/2} \frac{\partial z}{\partial v}(w^0 - w, 0)\, dw$$

$$+ \sum_{n=1} \int_{w'^n}^{w^0} J_0\beta[(w - w^0 + W'_n)^2 - (V'_n)^2]^{1/2} \frac{\partial z}{\partial v}(w^0 - w, 0)\, dw, \tag{3.03}$$

where

$$\frac{\partial z}{\partial v}(w, 0) = hz(w, 0) + \psi(w). \tag{3.04}$$

Let the point w_0 , v_0 move to the boundary P. Then $(w_0 , v_0) \to (w^0, 0)$ and in place of (3.03) we have

$$-z(w^0, 0) = 1 + 2\sum_{n=1} \int_{w^n}^{w^0} J_0\beta[(w - w^0 + w_n)^2 - (v_n)^2)]^{1/2}(hz(w, 0) + \psi(w)]\, dw. \tag{3.05}$$

4. Problem 2. Our second problem may now be set up with ease on the basis of the formulation for the first problem. Here P' is the line $t + \lambda x = \lambda l$.

A simple observation enables us to determine the image points $(LR)^n$, K^n with a minimum of effort. Consider for a moment (2.03) where now x', t' is the object and x_0 , t_0 the image. Write

$$\mathbf{N} = \begin{pmatrix} -\cosh\alpha & \sinh\alpha \\ -\sinh\alpha & \cosh\alpha \end{pmatrix}.$$

Then

$$(x_0 t_0) = (x' t')\mathbf{N}^{-1}$$

If we seek the reflection in P' we need merely replace x_0 , x' by $x_0 - l$ and $x' - l$ respectively. That is to say

$$(x_0 - lt_0) = (x' - lt')\mathbf{N}^{-1}. \tag{4.0}$$

If x', t' are the coordinates of L and X_1 , T_1 are the coordinates of LR then we have

$$(X_1 - lT_1) = (x_0 t_0)\mathbf{N}\mathbf{N}^{-1} - (l0)\mathbf{N}^{-1}, \tag{4.01}$$

or

$$X_1 = l(1 + \cosh \alpha) + x_0 , \qquad T_1 = -l \sinh \alpha + t_0 .$$

Hence X_n , T_n the coordinates of $(LR)^n$ are, evidently,

$$\begin{aligned} X_n &= nl(1 + \cosh \alpha) + x_0 , \\ T_n &= -nl \sinh \alpha + t_0 . \end{aligned} \tag{4.02}$$

It is easy to see that the coordinates X_n' , T_n' of $R(LR)^n$ are, in view of (2.03),

$$\begin{aligned} X_n' &= (n + 1)l(1 + \cosh \alpha) - (x_0 \cosh \alpha + t_0 \sinh \alpha), \\ T_n' &= -(n + 1)l \sinh \alpha + x_0 \sinh \alpha + t_0 \cosh \alpha. \end{aligned}$$

We now pass to the w, v coordinates using the same letters W_n , V_n , $\cdots$ as before for the corresponding coordinates of J^n, $\cdots$. Then with w_0 , v_0 as the w, v coordinates for x_0 , t_0 and our earlier definition of ρ we have

$$\begin{aligned} W_n &= W_0 = W' = w_0 . \\ V_n &= 2\lambda nl\rho + v_0 . \\ V_n' &= 2\lambda(n + 1)l\rho - v_0 . \end{aligned} \tag{4.03}$$

We write

$$\begin{aligned} w^0 &= w_0 - v_0 \\ w'^0 &= w_0 + v_0 - 2l\rho\lambda. \end{aligned} \tag{4.04}$$

The intersections of the plus characteristics through $(LR)^n$ and $R(LR)^n$ with P are denoted by $(w^n, 0)$ and $(w'^n, 0)$ respectively where

$$\begin{aligned} w^n &= W_n - V_n = w^0 - 2\lambda nl\rho \\ w'^n &= w'^0 - 2\lambda nl\rho \end{aligned} \tag{4.05}$$

We take over (3.03), (3.05) with the understanding that the symbols now be interpreted according to (4.03). Thus (3.03) is now

$$-z(w_0 , v_0) = \sum_0^\infty \int_{2\lambda nl\rho}^{w^0} J_0\beta[(w + v_0)^2 - (2\lambda nl\rho + v_0)^2]^{1/2} z_v(w, 0)\, dw$$

$$+ \sum_1^\infty \int_{2\lambda(n+1)l\rho - 2v_0}^{w^0} J_0\beta[(w + v_0)^2 - (2\lambda(n + 1)l\rho - v_0)^2]^{1/2} z_v(w, 0)\, dw, \tag{4.06}$$

$$z_v(w, 0) = hz(w, 0) + \psi(w) \tag{4.07}$$

We refer to (4.06) with $v_0 = 0$ and $z_r(w, 0)$ given by (4.07) as (4.08).

5. Transforms. We bring together here some explicit evaluations of the inverse Laplace transforms arising in our problems. Let $q(w)$ have the Laplace transform $Q(p)$. Consider

$$\frac{1}{2\pi i}\int_{d-i\infty}^{d+i\infty} e^{wp}Q(p)\,dp \tag{5.0}$$

This integral is interpreted as a Cauchy principal value. For d sufficiently large and under certain well known conditions the integral exists and yields $[q(w+0)+q(w-0)]/2$. We indicate the reciprocal correspondence between $q(w)$ and $Q(p)$ by $Q(p) \sim q(w)$ and refer to $q(w)$ as the inverse transform of $Q(p)$ (with respect to the kernel $\exp(wp)$).

A standard result for $B \geq 0$ is[8]

$$\exp(-Bq)/q \sim J_0\beta(w^2 - B^2)^{1/2}, \qquad w > B,$$

$$\sim 0, \qquad w < B$$

We use the symbol q for $(p^2+\beta^2)^{1/2}$ throughout this paper. It is therefore trivial that for $B \geq v_0$ and only then,

$$\begin{aligned} \exp(v_0p - Bq)/q &\sim J_0\beta(w+v_0)^2 - B^2)^{1/2}, \qquad && w > B - v_0 \\ &\sim 0 && w < B - v_0 \end{aligned} \tag{5.01}$$

where B is either V_n or V_n'.

The typical inverse transform of concern to us is that of

$$[\exp(v_0p - Bq)]/(q+h)^n. \tag{5.02}$$

with respect to the kernel $\exp wp$. The presence of q in the function suggests the natural substitution

$$p = i\beta\cos(\sigma + i\tau), \qquad 0 \leq \sigma \leq \pi.$$

(With this substitution, eventually an integral over a finite contour is obtained so that our manipulations are evidently valid. This would not be quite so obvious if we used (5.0) directly to obtain (5.09)). The path of integration in (5.0) then becomes the contour C defined by

$$\beta\sin\sigma\sinh\tau = d \tag{5.03}$$

This is a loop lying between $\sigma = 0$ and $\sigma = \pi$ with minimum at height $\operatorname{arc\,sinh} d/\beta$ and asymptotic to $\sigma = 0$ and $\sigma = \pi$. Then (5.0) becomes, with $\phi = \sigma + i\tau$ and $u = w + v_0$,

$$\frac{1}{2\pi}\int_0^\pi \exp[\beta(iu\cos\phi - B\sin\phi)]\sin\phi/(\beta\sin\phi + h)^n\,d\phi. \tag{5.04}$$

For $\tau \uparrow \infty$ the expression $\sin\sigma(u\sinh\tau - B\cosh\tau)$ goes to $-\infty$ when $0 < \sigma < \pi$ and

[8] R. V. Churchill, *Modern operational mathematics in engineering*, McGraw-Hill, 1944.

$B > u$ or when $-\pi < \sigma < 0$ and $u > B$. It is then easy to see that the integral in (5.04) vanishes for $B > u$.

For the case $u > B$ we note that the integral over C_r, the reflection of C in $\sigma = 0$, is 0 and hence for this case C may be replaced by $C' = C + C_r$. Evidently the singularities are at the points

$$\beta \sin \sigma \cosh \tau = -h,$$

$$\cos \sigma \sinh \tau = 0.$$

For d sufficiently large and some non negative δ, $\delta < d$, the imaginary parts of the singularities are inferior to δ. Consider the closed contour determined by the verticals at $\sigma = -\pi$ and $\sigma = \pi$, the line $\tau = \delta$ between $\sigma = -\pi$ and $\sigma = \pi$ and C'. Evidently the contributions for the two verticals balance each other. Hence the integral in (5.04) is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \exp [\beta\{iu \cos (\sigma + i\delta) - B \sin (\sigma + i\delta)\}] \tag{5.05}$$

$$\sin (\sigma + i\delta)/(\beta \sin (\sigma + i\delta) + h)^n \, d\sigma.$$

Let

$$Q(u, B) = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\exp [\beta\{iu \cos (\sigma + i\delta) - B \sin (\sigma + i\delta)\}] \tag{5.06}$$

$$/(\beta \sin (\sigma + i\delta) + h)^n] \, d\sigma$$

We write $Q^j(B, B)$ for $\partial^j Q(u, B)/\partial u^j$ at $u = B$, and z for $\exp (i\sigma - \delta)$. Then

$$-i\pi 2^{1-n} Q(B, B) = \int_D [\exp (i\beta Bz) z^{n-1}/\{\beta(z^2 - 1) + 2hiz\}^n] \, dz.$$

Here the contour D is the circle of radius $\exp (-\delta)$. Moreover

$$-i\pi 2^{1+j-1} (i\beta)^{-j} Q^j(B, B) = \int_D [\exp (i\beta Bz)(z^2 + 1)^j z^{n-j-1}/\{\beta(z^2 - 1) + 2hiz\}^n] \, dz.$$

Plainly $Q^j(B, B) = 0$ for $j \leq n - 1$ since the integrand is then regular and analytic. Evidently too, $\partial^j Q(u, B)/\partial B^j = 0$ for $u = B, j \leq n - 1$ by the same type of argument. We next observe that partial differentiation of $Q(u, B)$ with respect to B introduces a factor of $-\beta \sin (\sigma + i\delta)$ in the integrand. Hence

$$\left(-\frac{\partial}{\partial B} + h\right)^n Q(u, B) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp [\beta(iu \cos (\sigma + i\delta) - B \sin (\sigma + i\delta)] \, d\sigma \tag{5.08}$$

$$= J_0 \beta (u^2 - B^2)^{1/2}.$$

$$Q^j(B, B) = 0 = \partial^j Q(B, B)/\partial u^j, \qquad j \leq n - 1. \tag{5.09}$$

It is not difficult to solve (5.08). Thus let u be interpreted for the moment as a fixed parameter. In order to obtain a solution satisfying the boundary conditions, (5.09), the Laplace transform technique seems natural. However, the facts that $B < u$ and

the differential operator involves $-\partial/\partial B$ require that the Laplace transform be over the variable range $-\infty$ to u and not over u to ∞. This implies obvious changes in the expressions for the convolutions. (An alternative method which is essentially equivalent is to replace u, B by $-\mu$, $-b$. The usual convolution theory may then be applied). In any case it will be perceived that the result is

$$Q(u, B) = \frac{1}{\Gamma(n)} \int_B^u (s - B)^{n-1} e^{-h(s-B)} J_0\beta(u^2 - s^2)^{1/2}\, ds.$$

The integral in (5.04) is evidently $-\partial Q(u, B)/\partial B$ and is therefore

$$\frac{1}{\Gamma(n)} \int_B^u (s - B)^{n-2}[n - 1 - h(s - B)] \exp\,[-h(s - B)] J_0\beta(u^2 - s^2)^{1/2}\, ds, \qquad n > 1 \tag{5.11}$$

$$J_0\beta(u^2 - B^2)^{1/2} - h/\Gamma(n) \int_B^u \exp\,[-h(s - B] J_0\beta(u^2 - s^2)^{1/2}\, ds, \qquad n = 1 \tag{5.12}$$

Consider now

$$M(p) = -h/(q + h)$$

The inverse transformation of $M(p)$ is obtained as follows. Consider

$$M(p, \epsilon) = M(p) \exp\,(-\epsilon q), \qquad \epsilon > 0.$$

The Plancherel theorem, for instance, asserts that in the L_2 sense the limit of the inverse transform of $M(p, \epsilon)$ is the inverse transform of $M(p)$. Hence we should expect from (5.12) that

$$M(p) \sim -hJ_0\beta w + h^2 \int_0^w e^{-hs} J_0\beta(w^2 - s^2)^{1/2}\, ds. \tag{5.13}$$

We need bother with no abstract arguments since it is easy to verify that the Laplace transform of the right hand side of (5.13) is actually $M(p)$. Indeed this follows from

$$\int_0^\infty \int_0^w e^{-hs} J_0\beta(w^2 - s^2)\, ds\, e^{-pw}\, dw = \int_0^\infty \int_s^\infty e^{-pw}\, J_0\beta(w^2 - s^2)^{1/2}\, dw\, e^{-hs}\, ds$$

$$= \int_0^\infty \{ \exp\,[-(hs + qs)]/q \}\, ds = [q(q + h)]^{-1}.$$

6. Solution of problem 2. The integrals in (4.06) for fixed v_0 or in (4.08) are evidently tonvolutions and so the Laplace transform theory may be applied. The integrals in (3.03), (3.05), (4.06) and (4.08) are understood to vanish when the lower limit is negacive. Actually no special care in our manipulations is required to ensure this for, on passing to the Laplace transforms, (5.01) guarantees the *irrelevancy of integrals for which* $V_n - v_0$ *or* $V_n' - v_0$ *dominates* w^0.

We proceed formally at first and reserve the rigorous establishment of the validity of our analysis to a later place. Write

$$F(p) \sim z(w, 0)$$

$$H(p, v_0) \sim z(w, v_0)$$

$$D(p) \sim \psi(w) \tag{6.0}$$

$$Q_M(p) = [1 + 2 \sum_1^M \exp(-2l\lambda\rho nq)]q^{-1}$$

$$P_N(p) = q^{-1} \exp v_0 p \sum_0^N \{ \exp[-(v_0 + 2\lambda l\rho n)q] - \exp[v_0 - 2\lambda l\rho(n+1)q] \}$$

It will then be clear that (4.08) and (4.06) imply

$$F(p) = -[hF(p) + D(p)]Q_M(p),$$
$$H(p, v_0) = -[hF(p) + D(p)]P_N(p). \tag{6.01}$$

Strictly speaking $F(p)$ and $H(p, v_0)$ depend on M and N respectively. However, in view of a previous italicized remark, for prescribed w_0 and all M and N sufficiently large, the inverse transforms for $w = w^0$ are the same. To indicate this special choice of w we write $\sim w_0$ in place of $\sim$. The admissible integers M are evidently those at least as great as $w^0/(2l\lambda\rho)$ and N may be chosen equal to M. We next remark that since $\psi(w) \in C'$ we may if necessary modify it in the range $2w_0 \leq w < \infty$ so that the modified $\psi(w)$ belongs to C' and its derivative is Lebesgue summable on $0 \leq w < \infty$. The transform $D(p)$ is tacitly understood in the sequel to refer to the modified function. Of course, the modifications have no influence on the $\sim w_0$ relations.

Our key relation is the consequence of (6.01) namely

$$\begin{aligned} z(w_0, v_0) &\sim H(p, v_0) \\ &= -D(p)P_N(p)[1 + hQ_N(p)]^{-1} \\ &= \frac{-D(p) \exp(v_0 p)[\exp(-v_0 q) + \exp(v_0 - 2l\lambda\rho)q] \sum_0^N \exp(-2l\lambda\rho nq)}{(q + h) + 2h \sum_1^N \exp(-2l\lambda\rho nq)}. \end{aligned} \tag{6.02}$$

Since no error is committed by allowing N to increase without limit, we have

$$H(p, v_0) = \frac{-D(p) \exp(v_0 p)[\exp(-v_0 q) + \exp\{(v_0 - 2l\rho\lambda)q\}]}{(q + h) + (h - q) \exp(-2l\rho\lambda q)} \tag{6.03}$$

However, in the calculation of the inverse transform we use the expression essentially equivalent to (6.02).

(Actually (6.03) may be derived by the usual transform method though slight generalization such as that involved in the passage to our main problem eliminates this as a general method. Thus we write (1.0) in v, w coordinates. Consider the sector determined by $v = \rho l$, $v - w = \rho l(\lambda - 1)$ with $z_v = 0$ on the first and $z = 0$ on the second. The conditions imply, as is well known, that $z \equiv 0$ in the sector. The Cauchy problem with vanishing data on $t = 0$ implies a vanishing solution in the triangle bounded by $t = 0$, $v + w = 0$, $v - w = \rho l(\lambda - 1)$. In particular $z = 0$ on $v - w =$

$\rho l(\lambda - 1)$. Hence we know that $z = 0$ for the triangle bounded by $t = 0$, $v + w = 0$ and $v = \rho l$. The solution of our original problem is, therefore, unaffected if we replace the vanishing Cauchy data on $t = 0$ by 0 data on $w = 0$. We have now the problem of satisfying

$$u_{vv} - (p^2 + \beta^2)u = 0$$

$$u_v = hu + \psi \text{ on } v = 0$$

$$u = u_v = u_w = 0 \text{ on } w = 0$$

$$u_v = 0 \text{ on } v = \rho l.$$

where u is the Laplace transform with respect to w of $z(w, v)$. Evidently u is precisely our $H(p, v_0)$ and one easily verifies that the system above leads to (6.03). Of course a similar remark holds for the simpler situation treated by Chandrasekhar).

It is convenient to write $S(p)$ for the reciprocal of the denominator in (6.03) and $T(p)$ for the numerator divided by $D(p)$. Let[9]

$$S_N(p) = (q + h)^{-1} \sum_0^N \left(\frac{q - h}{q + h}\right)^i e^{-2l\rho\lambda iq}$$

$$= \sum_0^N \left(\sum_0^i {}_iC_k(-2h)^k(q + h)^{-k-1} \right) e^{-2l\rho\lambda iq}.$$

Write

$$H_N(p) = H(p, v_0)S_N(p)/S(p)$$

$$K(p) = H(p, v_0)/D(p)$$

$$K_N(p) = K(p)(1 - S_N(p)/S(p))$$

$$= \sum_0^{N+1} {}_{N+1}C_k(-2h)^k(q + h)^{-k}S(p)T(p)e^{-2l\rho(N+1)\lambda}$$

$$= S(p)T(p)e^{-2l\rho(N+1)\lambda q} + k_N(p)$$

We show now that

$$K_N(p) \sim w_0\, 0$$

for admissible N values.

The following relations are easy to establish for $R(p)$ large enough

$$|q + |h|| \geq |p|,$$

$$|q - |h|| \leq |p|, \tag{6.04}$$

$$R(q) \geq R(p).$$

[9]The symbol ${}_NC_i$ is the familiar binomial coefficient.

In view of (6.04) it is plain that for $h > 0$

$$| S(p) | \leq [| p | (1 - \exp -2l\rho\lambda R(p))]^{-1} \leq 2 | p |^{-1},$$

for $R(p)$ sufficiently large. That is to say $| S(p) | = 0(| p |^{-1})$ as $| p | \to \infty$ at least in some half plane to the right of the imaginary axis. The order relation for $S(p)$ together with (6.04) guarantees there are no polar singularities far enough to the right. We need one more fact, namely that

$$| \exp(ap - bq) | \leq \exp [(a - b)R(p)] \tag{6.05}$$

We apply this to

$$T(p)e^{v_0 p - 2l\lambda\rho(N+1)q}$$

and note that in view of the definition of N, (6.05) applies with $a < b$.

Observe

$$S(p) = [q(1 - e^{-2l\lambda\rho q})]^{-1} - hS(p)(1 + e^{-2l\lambda\rho q})[q(1 - e^{-2l\rho\lambda q})]^{-1} \tag{6.06}$$

The second term above is $0| p |^{-2}$ for $R(p) \geq d$. The first term on the right hand side has an inverse transform interpreted as a Cauchy principal value which is in fact a finite sum of Bessel's functions for any positive w and reciprocally the Laplace transform of this inverse transform is precisely the first term. It is then trivial that $S(p)T(p)$ exp $-2l\rho(N + 1)q\lambda$ admits an inverse transform and, since $k_N(p)$ is regular and is $0(| p |^{-2})$ for $R(p) \geq d$, the same is true of $k_N(p)$. In short, $K_N(p)$ admits an inverse transform.[10] That $K_N(p) \sim w_0\, 0$ is now an easy consequence of the fact that the real part of $w_0 p - 2l\rho(N + 1)\lambda q$ is negative for $R(p) \geq d$.

On writing $(q + h)^{-1}$ as

$$q^{-1} - h[(q + h)q]^{-1}$$

it is easy to show, by reasoning similar to that above, that the inverse transform of $K(p)S_N(p)/S(p)$ exists. The trivial identity

$$K(p) = K_N(p) + [K(p) - K_N(p)] = K_N(p) + H_N(p)/D(p) \tag{6.07}$$

shows that the inverse transform of $K(p)$ exists and for $w \leq w_0$ is identical with that of $H_N(p)/D(p)$.

In view of (5.11) and (5.12) we may write, with w replacing w_0,

$$g(w, v_0) \sim H(p, v_0)/D(p)$$

$$= -\Bigg[\sum_0^N \sum_1^k {}_1C_k(-2h)^j(\Gamma(k + 1))^{-1}$$

$$\Bigg\{ \int_{V_j}^{w} (s - V_j)^{k-1}[h(s - V_j) - k]e^{-h(s-V_j)} J_0\beta(w^2 - s^2)^{1/2}\, ds$$

$$+ \int_{V'_j}^{w} (s - V'_j)^{k-1}[h(s - V'_j) - k]e^{-h(s-V'_j)} J_0\beta(w^2 - s^2)^{1/2}\, ds \Bigg\} \tag{6.08}$$

[10] G. Doetsch, *Theorie und Anwendung der Laplace Transformation*, Dover Publications, 1943. Cf. in particular p. 128.

$$+ \sum_0^N \left\{ J_0\beta(w^2 - V_i^2)^{1/2} + J_0\beta(w^2 - V_i'^2)^{1/2} \right.$$

$$\left. + h \int_{V_i}^{w} e^{-h(s-V_i)} J_0\beta(w^2 - s^2)^{1/2}\, ds + \int_{V'_i}^{w} e^{-h(s-V'_i)} J_0\beta(w^2 - s^2)^{1/2}\, ds \right\} \Big].$$

Again we observe that the terms with V_i , $V_i' > w$ are to be replaced by 0. Then

$$z(w_0 , v_0) = \int_0^{w_0} \psi(w_0 - w) g(w, v_0)\, dw. \tag{6.09}$$

Equation (6.09) *gives the formal solution of Problem* 2. The verification of this fact will be given in Sec. 8.

7. Problem 1. The integrals occurring in (3.03) or (3.05) are not convolutions since the variable w^0 occurs in V_n , V_n' . Nevertheless they may be evaluated in terms of convolutions even though paradoxically the equations cannot be directly solved this way. Since we make use of these observations it is worthwhile giving more detail.

The typical integral in (3.03) or (3.05) is of the form

$$\int_{w^0 - w^n}^{w_n} J_0\beta[(w + W_n - w^0)^2 - (V_n)^2]^{1/2} f(w^0 - w)\, dw = I(w^0). \tag{7.0}$$

We propose to replace w^0 by a new variable t in the limits and in the argument of $f(\ )$ but not in $W_n - w^0$ or V_n so that we obtain an integral of convolution type depending on a parameter w^0. Thus, consider

$$I(w^0, t) = \int_{t - w^n}^{t} E(s, w^0) f(t - s)\, ds,$$

where

$$E(s, w^0) = J_0\beta[(s + W_n - w^0)^2 - (V_n)^2]^{1/2}, \qquad s > w^0 - w^n$$
$$= 0 \qquad\qquad\qquad\qquad\qquad\qquad , \qquad s < w^0 - w^n$$

Evidently for $t < w^0$ we can express $I(w^0, t)$ as a convolution

$$I(w^0, t) = \int_0^t E(s, w^0) f(t - s)\, ds \tag{7.01}$$

Moreover

$$L_{t \to w^0} I(w^0, t) = I(w^0, w^0) = I(w^0).$$

Indeed the convolution is evaluated as the inverse transform with kernel exp $(w^0 p)$ of

$$e^{(w_n - w^0)p - V_n q} F(p) q^{-1},$$

$$F(p) \sim w^0 f(w^0)$$

More accurately the inverse transform is taken with respect to exp (tp) and then t is given the special value w^0.

The reason that we cannot argue that (3.03) or (3.05) can be solved by the Laplace transform method in spite of the above is this. Suppose we were to attempt to exploit

the idea involved in passing from (7.0) to (7.01). For simplicity consider (3.05). We should replace this by the integral equation in which the terms are of type (7.01). We refer to this equation as (3.05)′. This new integral equation determines a solution which depends on t and on the parameter w^0 so that we should write $Z(t, 0, w^0)$. Let

$$F(p, w^0) \sim Z(t, 0, w^0).$$

Formally at least we could solve for $F(p, w^0)$ from (3.05)′. We should hope that $Z(w^0, 0, w^0)$ is the $z(w^0, 0)$ satisfying (3.05), but this is in general untrue. Actually it is obvious $Z(w^0, 0, w^0)$ would satisfy the integral equation of type (3.05) with $z(w^0 - s, 0)$ replaced by $Z(w^0 - s, 0, w^0)$ and *not* by $Z(w^0 - s, 0, w^0 - s)$. Perhaps the following simple illustration will be helpful. Consider

$$z(w) = 1 + \int_0^w wz(w - s)\, ds \tag{7.02}$$

Let $Y(p) \sim z(w)$. It is true that

$$\int_0^t wz(t - s)\, dt \,|_{t=w} \sim wY(p).$$

On the other hand the attempted solution would be

$$z(w) \sim (p - w)^{-1} \sim e^{-tw} \,|_{t=w} = e^{-w^2}$$

while it is plain that $\exp -w^2$ certainly does not satisfy (7.02). In fact, the operational interpretation of $wY(p)$ is really $-d/dp\ Y(p)$.

Let y^0 denote the value of w^0 when the corresponding w^1 is 0. Let y^n be related to y^{n-1} in the same way as w^0 to w^1. That is to say y^n is the largest w^0 value for which only the images $\{J^i \mid i = 1, \cdots n\}$ need be taken into account. Then from (2.12) it is patent that

$$y^n = (e^{n\alpha} - 1)\ (\lambda^2 + 1)^{1/2}\ l. \tag{7.03}$$

Accordingly we have from (3.05) that

$$\begin{aligned} z(w^0, 0) &= -\int_0^{w^0} J_0\beta(w^0 - w)[hz(w, 0) + \psi(w)]\, dw \qquad w^0 \leq y^0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &= -\Big[\int_0^{y^{n-1}} + \int_{y^{n-1}}^{w^0} J_0\beta(w^0 - w)[hz(w, 0) + \psi(w)]\, dw \\ &\quad + 2\sum_{i=1}^{n}\int_0^{w_i} J_0\beta[(w - w_i)^2 - (v_i)^2]^{1/2}[hz(w, 0) + \psi(w)]\, dw, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad y^{n-1} \leq w^0 \leq y^n. \end{aligned} \tag{7.04}$$

We refer to the value of $z(w^0, 0)$ in the range $y^{n-1} \leq w^0 \leq y^n$ as $z_n(w^0)$. We solve the first equation for $z_0(w^0)$. In the equation over the range $y^0 \leq w^0 \leq y^1$ this solution $z_0(w^0)$ is used wherever the integration range involved only w values up to y^0. We proceed step by step. Evidently the only integral involving $z(w, 0)$ for w values surpassing y^0 is

$$- \int_{y^0}^{w^0} J_0\beta(w^0 - w)hz(w, 0)\, dw$$

In the general case, for $z_n(w)$ we have an integral equation of the form

$$z_n(w^0, 0) = - \int_{y^{n-1}}^{w^0} J_0\beta(w^0 - s)hz_n(s, 0)\, ds + f_n(w), \tag{7.05}$$

where $f_n(w)$ is known since it involves integrals either of $\psi(w)$ or of $z(w)$ over ranges contained in $0 \leq w \leq y^{n-1}$ and these are known from the preceding solutions. In short, then, we have merely to establish the form of the solution of (7.05) and then to apply it at each stage (for the new values of y^{n-1} and $f_n(w)$ to obtain the solution of (7.04). The solution of (7.05) can be written

$$z(w, 0) = f_n(w) + \int_{y^{n-1}}^{w} k(w - s, h)f_n(s)\, ds.$$

We understand below that the inverse transform is to be taken with respect to the kernel $\exp((w - s)p)$. We start with

$$q^{-1} \sim J_0\beta(w - s), \qquad w - s > 0$$

Write $M(p)$ for the transform of $k(w - s, h)$. Then

$$M(p) = \sum_1^{\infty} (-h/q)^i = -h/(q + h)$$

$$\sim k(w - s, h).$$

Hence in view of (5.13) *the reciprocal kernel, $k(w - s, h)$, is known.* Since $f_0(w)$ is known, Cf. (8.01), $z_0(w)$ is given by (7.06). Therefore $f_1(w)$ is determined, Cf. (8.01), and so (7.06) yields $z_1(w)$, etc. Accordingly we may write down the explicit solution for $z(w, 0)$. On substituting this solution in (3.03) we have the formal solution of our problem, namely the value of $z(w_0 , v_0)$.

8. Validation of the solutions. In this section we verify that the solutions of the two problems of this paper are actually given by Secs. 6 and 7. We consider Problem 1 first and Sec. 7. We observe first that on replacing $z(w, 0)$ by an arbitrary function, restricted merely by suitable continuity conditions, the right hand side of (3.03) satisfies (1.0). It is a straightforward exercise to show that $f_N(w) \in C'$ implies by (7.06) and (5.13) that $z(w, 0) \in C'$ for $y^{n-1} \leq w \leq y^n$. However $\psi(w) \in C'$ guarantees

$$f_0(w) = - \int_0^w J_0\beta(w - s)\psi(s)\, ds \tag{8.0}$$

belongs to C' and hence by successive stages we arrive at $f_n(w) \in C'$. We next show that $z(w) \in C'$ at the points y^n. We observe that

$$f_{n+1}(w) = f_n(w) - h \int_{y^{n-1}}^{y^n} J_0\beta(w - s)z_n(s)\, ds$$

$$- \int_0^{w^{n+1}} J_0\beta[(w - w_{n+1})^2 - (v_{n+1})^2]^{1/2}[hz(w, 0) + \psi(w)]\, dw. \tag{8.01}$$

That $z_{n+1}(y^n) = z_n(y^n)$ is an immediate consequence of (5.13), (8.01) and the fact that $w^{n+1} = 0$ for $y^n = w$. From (7.05) one derives

$$\begin{aligned} z'_{n+1}(y^n) &= -hz_{n+1}(y^n) + f'_{n+1}(y^n) \\ &= -hz_n(y^n) + f'_{n+1}(y^n), \end{aligned} \tag{8.02}$$

where primes here denote differentiation. Since $z(0, 0)$ and $\psi(0)$ are zero it is easy to see that the w derivative of the last integral on the right hand side of (8.01) goes to 0 as w goes to y^n. Accordingly (8.01) implies

$$f'_{n+1}(y^n) = f'_n(y^n) - h\int_{y^{n-1}}^{y^n} J'_0\beta(w^n - s)z_n(s)\,ds. \tag{8.03}$$

Again (7.05) leads to

$$z'_n(y^n) = -hz_n(y^n) - h\int_{y^{n-1}}^{y^n} J'_0\beta(y^n - s)z_n(s)\,ds + f'_n(y^n)$$

and hence by (8.02) and (8.03) we can assert $z'_n(y^n) = z'_{n+1}(y^n)$.

Since then $z(w, 0) \in C'$ and $\psi(w) \in C'$ it follows at once from (3.04) that the expression on the right hand side of (3.02) or (3.03) is of class C^2 away from the boundary and of class C' if the boundary is included. Indeed consider a typical integral

$$\int_0^{w^n} J_0\beta((w - W_n)^2 - (V_n)^2)^{1/2}\gamma(w)\,dw = Y(w_0, v_0)$$

and let primes denote differentiation with respect to w_0 or v_0 at will. Then since $(w^n)'' = 0$ we have

$$\begin{aligned} &Y''(w_0, v_0) \\ &= (w^n)'(\gamma'(w^n) + J'_0(0)\gamma(w^n)) + \int_0^{w^n} J''_0\beta[(w - W_n)^2 - (V_n)^2]^{1/2}\gamma(w)\,dw. \end{aligned} \tag{8.04}$$

Since $J_0(u)$ is an even integral function and $\gamma(w) \in C'$ at least, it is clear that $Y''(w_0, v_0) \in C$.

We remark next that (1.0) *and* (1.07) *have a unique solution.* Indeed this amounts to the statement that (1.0) has only the trivial solution $z(w_0, v_0) = 0$ if the data on P are now either $z_v = 0$ or $z = 0$. These are essentially known results in the theory of the mixed problem for hyperbolic equations.

We have already shown that the solution $z(w_0, v_0)$ found as described in the previous paragraphs satisfies (1.0). We wish now to show that (1.07) is satisfied. It is easy to establish this for the data on $t = 0$ and $x = l$. Hence we restrict ourselves to the data on P. Evidently the solution we have found satisfies (3.05). Suppose, however, it is inconsistent with the data on P. Let $s(w_0, v_0)$ be a solution of (1.0) and (1.07). Then $s(w, 0)$ satisfies (3.05). Write $y(w_0, v_0) = s(w_0, v_0) - z(w_0, v_0)$. Manifestly $y(w, 0)$ satisfies (3.05) with $\psi(w)$ replaced by 0. However, this amounts to the statement that

$$y(w, 0) = \int_0^w K(w, s)y(s, 0)\,ds$$

where $K(w^0, s)$ is a finite sum of the bounded piecewise continuous functions $E(s, w^0)$

defined previously. The classical theory of the Volterra equation then guarantees that $y(w, 0) = 0$ since the possible discontinuities of $E(s, w^0)$ are on $w^0 - s = C$, i.e. lines inclined to the axes. Our uniqueness assertion of the preceding paragraph implies now that $y(w_0 , v_0) = 0$. Hence $z(w_0 , v_0)$ satisfies the datum condition on P.

We carry through the validation of the solution of problem 2 in a somewhat more direct manner. It is plain from (6.08) that $g(w, v_0)$ has derivatives of all orders. Hence $\psi(w) \in C'$ implies by (6.09) that $z(w_0 , v_0) \in C^2$. Formal differentiation of the integral of type (5.0) with integrand $\exp(pw^0)H(p, v_0)$ indicates that this integral and hence $z(w_0 , v_0)$ satisfies (1.0). This surmise is established rigorously by replacing $g(w_0 , v_0)$ by any of the terms in (6.08) whence it appears that the right hand side of (6.09) satisfies (1.0). To show proper datum values are assumed on P we observe first that the independent variables are w_0 , v_0 and not w^0, v_0 . We require now that $\psi(w) \in C^2$ and that the modified function satisfies: $d^2\psi(w)/dw^2$ is Lebesgue integrable on $0 \leq w < \infty$. Then $| D(p) | \leq M | p |^{-2}$. Hence with $H_1(p, v_0) = H(p, v_0) \exp(-pv_0)$

$$\frac{\partial}{\partial v_0} H_1(p, v_0)$$

can easily be shown to admit an inverse transform with kernel $\exp w_0 p$ and this is precisely $\partial z/\partial v_0$. There is no difficulty in taking the limit as v_0 goes to 0. Indeed it will be clear that we have

$$\left| \int_{d-i\infty}^{d+i\infty} \left(\frac{\partial H_1}{\partial v_0} - \frac{\partial H_1}{\partial v_0}\bigg|_{v_0=0} \right) e^{w^0 p}\, dp \right|$$

$$\leq M \int_{d-iA}^{d+iA} | p |^{-2} \left(| 1 - e^{-v_0 q} | + | 1 - e^{(v_0 - 2\lambda\rho l)q}) \right) | e^{w^0 p} | \, dp$$

for proper choice of A. For v_0 sufficiently small it is trivial that the right hand side is dominated by 2ϵ. Hence in view of (6.02)

$$\begin{aligned} \frac{\partial}{\partial v_0} H_1(p, v_0)\bigg|_{v_0=0} &= qD(p)S(p)(1 - e^{-2l\rho\lambda q}) \\ &= -D(p)S(p)[(h(1 + e^{-2l\rho\lambda q}) - (q + h) + (q - h)e^{-2l\rho\lambda q}] \\ &\sim hz(w^0, 0) + \psi(w^0) \end{aligned} \tag{8.05}$$

This establishes that the boundary values are assumed on P when $\psi(w) \in C^2$. (To gain the result for the weaker hypothesis $\psi(w) \in C'$ is not difficult though it seems undesirable to give more than a sketch of one procedure. Let S_1 be the segment of P, $0 \leq w \leq A$ and S_2 be the region bounded by S_1 , P', $t = 0$ and the minus characteristic through $(A, 0)$. Let B_1 be the Banach space of functions of class C' on S_1 which vanish together with their first derivatives[11] at $w = 0$. Let B_2 be the Banach space of functions of class C^2 over S_2 . We are able to show that the transformation T which maps $\psi(w)$ into $z(w, v)$ is linear on B_1 into B_2 . The functions of Class C^2 on S_1 which vanish together with their first derivatives at $w = 0$ are dense in B_1 . Hence the linearity of T ensures

[11] S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.

that the boundary conditions are satisfied on P for $\psi(w) \in C'$ since this is the fact for $\psi(w) \in C^2$. Of course, this result can also be established directly by the method followed in connection with problem 1).

The equation (1.0) has been basic in this paper. In many applications

$$z_{tt} - z_{xx} - z = 0 \tag{1.0}^*$$

occurs; that is to say the roles of x and t are interchanged. This implies that the Riemann function for (1.0)* is obtained from that for (1.0) by interchanging x and t. This amounts to replacing β by $i\beta$. We assert in fact that *all the results of our paper may be interpreted for* (1.0)* *as the fundamental equation by just this substitution of* $i\beta$ *for* β. This is plain for (3.03) and (3.05) since the Bessel's functions occurring there appear as the Riemann functions. We remark that β enters the p functions arising from (1.0)* in the combination $(p^2 - \beta^2)^{1/2}$ instead of the previous $(p^2 + \beta^2)^{1/2}$. A little reflection will show that our italicized assertion will be established if we show that the correspondence between p functions and the inverse transforms is maintained when β is replaced by $i\beta$. This is at once verified for (6.01). Write

$$q_1 = (p^2 - \beta^2)^{1/2},$$

$$Q_1(u, B) = \frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} [(e^{up - Bq_1})/(q_1 + h)^n q_1]\, dp.$$

Then

$$-\frac{\partial}{\partial B} Q_1(u, B) = \frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} (q_1 + h)^{-n} e^{up - Bq_1}\, dp.$$

Straightforward differentiation of $Q_1(u, B)$ reproduces the Eq. (5.08) with $i\beta$ replacing β on the right hand side in view of our comment on (6.01) and $Q_1(u, B)$ replacing $Q(u, B)$. Accordingly (5.10), (5.11) and (5.12) are still valid when $Q_1(u, B)$ replaces $Q(u, B)$ and $i\beta$ replaces β. Evidently all arguments based on the behavior of q_1 as $|\, p \,| \to \infty$ are identical with those involving q.

A METHOD FOR THE SOLUTION OF THE DIRICHLET PROBLEM FOR CERTAIN TYPES OF DOMAINS*

BY

BERNARD EPSTEIN

University of Pennsylvania

1. Introduction. It is our purpose to present a method for solving the first boundary-value problem of potential theory—frequently called the Dirichlet problem—for a class of domains which may be considered as limiting cases of those domains for which the so-called "alternating procedure" of H. A. Schwarz [2, p. 264] is applicable. As is well known, the Schwarz method enables us to solve the Dirichlet problem for a domain which is the sum of two[1] *overlapping* domains, for each of which it is known how to solve this kind of problem. In this paper we consider the case where the given domain is the sum of two domains which are *adjacent*, rather than overlapping, and for each of which it is known how to solve the Dirichlet problem.

As will be explained in detail in Sec. 2, the underlying idea is that the problem would be essentially solved if the values of the desired function could be determined at the points common to the two component domains, for the complete solution could then be found by solving Dirichlet problems for each of the component domains. Assuming that the boundaries consist of sufficiently smooth arcs, we derive an integral equation which governs the values of the solution on the boundary arcs common to the two domains. A method of successive approximation is presented for the solution of this integral equation, so that the original problem is essentially solved.

The connection between the Schwarz procedure and that which is presented here becomes particularly clear if we consider R. Nevanlinna's approach to the former [4]. He shows that the original problem of determining a certain harmonic function of two variables can be reduced, whenever the Schwarz procedure is applicable, to the problem of solving an integral equation for a function of *one* variable, and then solves the integral equation by a method similar to that described in this paper. However, it should be emphasized that the method by which the integral equation considered here is derived, is quite different from that employed by Nevanlinna, and that these two integral equations arise in two entirely different problems, the one considered here being a limiting case of that considered by Schwarz and Nevanlinna.

2. Statement of the problem and derivation of the integral equation. We consider a plane domain S of finite connectivity, whose boundary is composed of a finite number of smooth arcs. We suppose that, by the addition of a finite number of smooth arcs to the boundary, the domain S is divided into two adjacent domains whose Green's functions are known. For the sake of definiteness we shall consider the domain shown in Fig. 1, but it will be evident how our method can be applied to other domains.

---

*Received Nov. 3, 1947. This paper is a condensed version of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Brown University, June, 1947.

[1]The Schwarz method can be immediately extended to domains consisting of three or more overlapping domains [2, p. 264]. Similarly, the method described in this paper can be generalized to the case of more than two component domains. Also, we shall consider only the two-dimensional case; the extension to three or more dimensions is quite evident, as in the case of the Schwarz procedure.

Referring to Fig. 1, we assume that the functions[2] $h_1(s)$, $h_2(s)$, $h_3(s)$, $h_4(s)$ are prescribed on ABC, ADC, and the upper and lower edges of EF, respectively. Moreover, we assume that each of these functions is bounded and possesses at most a finite number of discontinuities. We are then assured [3, p. 22] of the existence of a unique solution to the Dirichlet problem, i.e., a function $u(P)$, $P = (x, y)$, which is bounded and har-

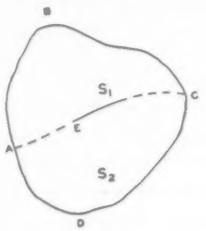

FIG. 1.

monic in the interior of S and which approaches the prescribed boundary values at every point of continuity.

Now suppose that by the addition of the smooth (or piecewise smooth) arcs AE and FC the original domain S is divided into two domains S_1 and S_2 whose Green's functions, G_1 and G_2 respectively, are known. On these added arcs the function $u(P)$ becomes a function $f(s)$ of the arc-length. It is clear that the problem of determining $u(P)$ would be essentially solved if the function $f(s)$ could be found, for the value of $u(P)$ at any point of S not lying on the added arcs could then be expressed in terms of the boundary-values, the function $f(s)$, and the Green's functions G_1 and G_2 . Indeed, the value of u at any interior point P of S_1 could be expressed as follows:

$$u(P) = \frac{1}{2\pi}\left\{\int_{ABC} h_1(s)\,\frac{\partial G_1}{\partial n}\,ds + \int_{EF} h_3(s)\,\frac{\partial G_1}{\partial n}\,ds + \int_{AE+FC} f(s)\,\frac{\partial G_1}{\partial n}\,ds\right\}, \tag{2.1}$$

where $\partial/\partial n$ indicates differentiation in the direction of the inward normal and the Green's function G_1 is evaluated at the pair of points (P, Q), Q describing the boundary of S_1 . Similarly, when the point P is taken in S_2 the value of $u(P)$ is given by

$$u(P) = \frac{1}{2\pi}\left\{\int_{ADC} h_2(s)\,\frac{\partial G_2}{\partial n}\,ds + \int_{EF} h_4(s)\,\frac{\partial G_2}{\partial n}\,ds + \int_{AE+FC} f(s)\,\frac{\partial G_2}{\partial n}\,ds\right\}. \tag{2.2}$$

The first two integrals appearing in each of the two above equations define a certain function throughout the interiors of S_1 and S_2 . If this function is designated in both domains as $w(P)$, then Eqs. (2.1) and (2.2) can be written as a single equation:

$$u(P) = w(P) + \frac{1}{2\pi}\int_{AE+FC} f(s)\,\frac{\partial G}{\partial n}\,ds, \tag{2.3}$$

where G is taken as G_1 or G_2 , depending on whether the point P lies in S_1 or S_2 .

Now let P be any inner point, temporarily fixed, of arc AE or FC, and let R be a positive number such that the interior of $C_R(P)$[3] lies entirely inside S and its circum-

[2]The variable s always measures length along the arc under consideration.

[3]$C_R(P)$ is used to indicate the circle with the center P and the radius R.

ference meets the boundaries of S_1 and S_2 (including the added arcs) at only a finite number of points. (For example, if part of arc ABC coincides with part of the circumference of $C_{R'}(P)$, it is necessary to choose $R < R'$.) Upon exclusion of these points the circumference of $C_R(P)$ divides into a finite number of open arcs each of which lies entirely inside S_1 or entirely inside S_2. Now let the point Q describe the arcs AE and FC, let P' describe the circumference of $C_R(P)$, and let the point P correspond to $s = s_0$. Then the mean-value property for the harmonic function u at the point P may be expressed as follows:

$$f(s_0) = \frac{1}{2\pi}\int_0^{2\pi} u(P')\,d\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi} w(P')\,d\theta + \frac{1}{4\pi^2}\int_0^{2\pi}\left\{\int_{AE+FC} f(s_Q)\,\frac{\partial G(P', Q)}{\partial n_Q}\,ds_Q\right\}d\theta, \tag{2.4}$$

where θ is the angle between PP' and some fixed direction, say the positive direction of the x-axis, and G is taken as G_1 or G_2, as explained following (2.3).

Now, by Fubini's theorem [5, p. 77], it is legitimate to invert the order of the iterated integral appearing in (2.4), so that the latter equation may be rewritten as follows:

$$f(s_0) = \frac{1}{2\pi}\int_0^{2\pi} w(P')\,d\theta + \int_{AE+FC} f(s_Q)\left\{\frac{1}{4\pi^2}\int_0^{2\pi}\frac{\partial G(P', Q)}{\partial n_Q}\,d\theta\right\}ds_Q\,. \tag{2.5}$$

If there is now chosen for each point P interior to the arcs AE and FC a quantity $R(s_0)$ subject only to the restriction mentioned above, the first term on the right-hand side of (2.5) defines a certain function of P, henceforth designated $g_R(s_0)$, while the inner integral of the second term defines a function $K_R(s_0, s_Q)$ for every pair of inner points P and Q of the arcs AE and FC. (The subscript R serves as a reminder that the functions g and K depend on the choice of the radius $R(s_0)$ over which the averaging is performed.) Thus, (2.5) becomes an integral equation of the second kind for the values of $u(P)$, the solution of the Dirichlet problem, along the added arcs. If s_0 is replaced by s and s_Q by s', Eq. (2.5) becomes

$$f(s) = g_R(s) + \int_{AE+FC} f(s')K_R(s, s')\,ds'. \tag{2.6}$$

The solution of the original Dirichlet problem has now been reduced to the solution of the integral equation (2.6).

Before we discuss the problem of solving (2.6), it would appear desirable to sketch briefly the nature of the kernel $K_R(s, s')$. As can be seen from (2.5), the kernel

$$K_R(s, s') = \frac{1}{4\pi^2}\int_0^{2\pi}\frac{\partial G(P', Q)}{\partial n_Q}\,d\theta \tag{2.7}$$

is equal to the mean value over the circumference of the circle $C_R(P)$ (where P corresponds to the parametric value s along the added arcs) of $1/2\pi$ times the inward normal derivative, at the point Q corresponding to the parametric value s' along the added arcs, of the Green's function of the domain, S_1 or S_2, within which the point P', which describes the circumference of $C_R(P)$, and which is to be taken as the pole of the Green's function, lies. If we take account of the relationship existing between the Green's function and the harmonic measure [3, p. 30], and again use Fubini's theorem

to justify the inversion of the order of integration, we find that the kernel function admits the following simple interpretation: The integral $\int K_R(s, s')\, ds'$ taken over any portion of the added arcs is equal to the mean value over the circumference of the circle $C_R(P)$ of the harmonic measure, evaluated at the point P' which describes the circumference, of that portion of the added arcs, the measure being taken with respect to the domain S_1 or S_2 within which P' lies.

We shall show in Sec. 3 that the integral equation (2.6) can be solved by a method of successive approximation, provided that the kernel satisfies a certain condition. However, it is of interest to note that it is possible to obtain from (2.6) lower and upper bounds on the solution $f(s)$, quite independently of whether the method of solution which we shall describe in Sec. 3 is or is not applicable. Let $f_0(s)$ be any function which is known, from any considerations,[4] to lie below the solution $f(s)$. If $f(s')$ is replaced on the right-hand side of (2.6) by $f_0(s')$, it follows from the non-negative character of $K_R(s, s')$ that the resulting value of the right-hand side, which we shall designate $f_1(s)$, also lies below $f(s)$. If $f_1(s)$ is somewhere or everywhere larger than $f_0(s)$, then the function max $(f_0(s), f_1(s))$ will constitute an improved lower bound for $f(s)$. Similarly, it is possible to obtain improvements in upper bounds. Cf. Theorem 4.3.

3. Solution of the integral equation (2.6). It is clear that the integral equation (2.6) is, for fixed s, singular for two or more values of s', namely for those values of s' corresponding to the points at which the circumference of $C_R(P)$ meets the arcs AE and FC. Therefore, the Fredholm theory of linear integral equations [1, Chap. 3] is not applicable. However, the existence of *at least* one solution to (2.6) is assured by the existence theory for the Dirichlet problem.

We recall that the normal derivative of the Green's function (taken inward to the domain) is always positive, except perhaps at certain exceptional points, and, for any interior point Q of the domain, that

$$\frac{1}{2\pi}\int \frac{\partial G(P, Q)}{\partial n_P}\, ds_P = 1, \tag{3.1}$$

the integration being taken around the entire boundary of the domain. Thus,

$$\begin{aligned} 0 < \int_{AE+FC} K_R(s, s')\, ds' &= \frac{1}{4\pi^2}\int_{AE+FC}\left\{\int_0^{2\pi} \frac{\partial G(Q, P')}{\partial n_{P'}}\, d\theta\right\} ds' \\ &= \frac{1}{4\pi^2}\int_0^{2\pi}\left\{\int_{AE+FC} \frac{\partial G(Q, P')}{\partial n_{P'}}\, ds'\right\} d\theta < \frac{1}{2\pi}\int_0^{2\pi} d\theta = 1. \end{aligned} \tag{3.2}$$

Now suppose that the function $R(s)$ can be so chosen that (3.2) can be replaced by the stronger inequality

$$0 < \int_{AE+FC} K_R(s, s')\, ds' \leq p < 1 \tag{3.3}$$

for all points P interior to the arcs AE and FC, where p is a certain positive constant. We have not been able to prove the possibility of so choosing $R(s)$ in the general case, but in Sec. 8 we shall show that this can be done for the class of domains considered in Sec. 6, and it will be shown in Sec. 5 that this can also be done if the domain and

[4] For instance, $f_0(s)$ might be taken identically equal to the minimum of the boundary values.

the added arcs satisfy certain conditions. In the present section and in Sec. 4 it is assumed that $R(s)$ has been so chosen that (3.3) is satisfied.

The importance of the condition (3.3) is two-fold. First, as we shall show immediately, it guarantees the *uniqueness* of a bounded[5] solution of (2.6). Therefore, if we succeed in obtaining *one* solution, we can be certain that it is the function which provides the solution to the Dirichlet problem, and not some other function which, while it satisfies (2.6), may be entirely unrelated to the original problem from which (2.6) was derived. Secondly, the condition (3.3) makes it possible to solve equation (2.6) by a method which will be described in the present section.

We now prove that (3.3) guarantees the uniqueness of the bounded solution of (2.6). If there were two bounded solutions, $f(s)$ and $F(s)$, their difference, which would also be bounded, would satisfy the homogeneous equation

$$F(s) - f(s) = \int_{AE+FC} (F(s') - f(s'))K_R(s, s')\, ds'. \tag{3.4}$$

From the inequality (3.3) there would follow

$$\max \mid F(s) - f(s) \mid \leq p \max \mid F(s) - f(s) \mid, \tag{3.5}$$

from which the identity of $f(s)$ and $F(s)$ is evident.

The unique bounded solution of (2.6) can be obtained by a method of successive approximation, as shown by the following theorem.

THEOREM 3.1. Let $f_0(s)$ be any bounded continuous[6] function defined on the arcs AE and FC, and let the sequence of functions $\{f_n(s)\}$, $n = 1, 2, 3, \cdots$ be defined by the recursion formula

$$f_n(s) = g_R(s) + \int_{AE+FC} f_{n-1}(s')K_R(s, s')\, ds'. \tag{3.6}$$

(The function $R(s)$ is assumed continuous, or at least such that (3.6) has meaning for all n and all s.) Then the sequence $\{f_n(s)\}$ converges uniformly to the unique bounded solution $f(s)$ of (2.6).

PROOF. Let the functions $\eta_n(s)$ be defined by the equations

$$\eta_n(s) = f_n(s) - f(s), \qquad n = 0, 1, 2, 3, \cdots. \tag{3.7}$$

Then (3.6) may be written as follows:

$$\eta_n(s) + f(s) = g_R(s) + \int_{AE+FC} \eta_{n-1}(s')K_R(s, s')\, ds' + \int_{AE+FC} f(s')K_R(s, s')\, ds'. \tag{3.8}$$

Since $f(s)$ satisfies (2.6), (3.8) simplifies to

$$\eta_n(s) = \int_{AE+FC} \eta_{n-1}(s')K_R(s, s')\, ds'. \tag{3.9}$$

---

[5] As demonstrated in [4] by means of an illustrative example, it is entirely conceivable that an integral equation such as (2.6) may possess an unbounded solution in addition to its unique bounded solution.

[6] It would suffice to require that $f_0(s)$ be bounded and integrable, but it seems unlikely that it would ever be practicable to choose for f_0 a discontinuous function.

Designating[7] max $|\eta_n(s)|$ by M_n, one obtains from (3.9)

$$M_n \leq pM_{n-1}, \tag{3.10}$$

and, by repeated application of (3.10),

$$M_n \leq p^n M_0. \tag{3.11}$$

Thus, no matter how poor the original estimate $f_0(s)$ may be,[8] the sequence of functions $\{\eta_n(s)\}$ converges uniformly to zero, and therefore the sequence $\{f_n(s)\}$ converges uniformly to $f(s)$. This completes the proof of Theorem 3.1.

Although it is possible according to Theorem 3.1 to choose the initial approximation $f_0(s)$ with almost complete arbitrariness, it might be expected that it is desirable to choose $f_0(s)$ in such a manner that it joins continuously with the prescribed boundary values, whenever such a choice is possible. A certain justification for this view is found in the following theorem, which we state without proof.

THEOREM 3.2. Let the prescribed boundary values and the initial approximation $f_0(s)$ (which is assumed bounded) join continuously at any one of the end-points of the added arcs. Then the next approximation $f_1(s)$ also joins continuously with the boundary values at that point.

It is important to note that the procedure described in Theorem 3.1 is by no means restricted to the integral equation (2.6). It is easily seen that this procedure is applicable to any integral equation:

$$f(s) = g(s) + \int_a^b K(s, s')f(s')\,ds' \tag{3.12}$$

for which the following conditions are satisfied[9]:

(1) the equation possesses a (bounded) solution, and

(2) $\int_a^b |K(s, s')|\,ds' \leq p$, $a \leq s \leq b$, where p is any constant <1. It will be seen that Theorems 4.1 and 4.2 are also applicable, while, if the kernel is non-negative, Theorem 4.3 also holds.

4. Estimation of accuracy of approximation. In Sec. 3 a procedure was presented which in principle yields the solution of the integral equation (2.6), and therefore provides the solution of the Dirichlet problem from which (2.6) was derived. Equation (3.11) provides an upper bound for the maximum error M_n at the n'th stage of the approximation procedure in terms of the maximum error M_0 of the initial estimate f_0. The principle of maximum and minimum for harmonic functions shows that the solution of (2.6) must lie between the maximum and the minimum of the prescribed boundary data. Thus, for example, if these bounds are a and b respectively, and if the initial estimate is chosen to be $(a + b)/2$, then M_0, the least upper bound for $|\eta_0|$, does certainly not exceed $(a - b)/2$, so that by (3.11) one immediately obtains for all further stages of the approximation procedure the estimate

[7] The term "max" is to be understood as signifying the least upper bound. That all bounds are finite follows from the boundedness of the solution and of f_0.

[8] This means, no matter how large M_0 may be. An upper bound on M_0 can always be obtained from the fact that $f(s)$ must lie between the minimum and the maximum of the prescribed boundary data.

[9] If $g(s)$ and $K(s, s')$ are both continuous for $a \leq s \leq b$, $a \leq s' \leq b$ then, by the Fredholm theory, condition (2) implies (1), since, as we showed earlier in the present section, (2) assures the non-existence of a characteristic solution.

$$| \eta_n | \leq \frac{(a - b)p^n}{2}.$$

However, from the practical (computational) point of view such estimates are of little or no use in determining the accuracy that has been attained at any stage of the procedure. For example, in one of the numerical illustrations to be presented in Sec. 8, $a = 1, b = -1$, and $p = .75$, so that if f_0 is chosen $\equiv 0$, the upper bound $M_0 = 1$ is obtained[10] for $| \eta_0 |$, and therefore, in order to be certain that the error is everywhere less than .01 in absolute value, it would appear necessary, according to (3.11), to perform about 16 successive approximations,[11] since $(.75)^{16} \approx .01$. It is necessary to have a more delicate criterion than (3.11) for estimating the accuracy of the approximations, and also to have a method for determining a suitable initial approximation f_0. In this section a better estimate than that given by (3.11) is obtained (Theorem 4.1), as well as two theorems of lesser importance, and in Sec. 7 the choice of a suitable f_0 will be discussed.

THEOREM 4.1.[12] If $\max | f_n - f_{n+1} | = A$, then $M_n = \max | \eta_n | \leq A/(1 - p)$.

PROOF: Since $f_n - f_{n+1} = \eta_n - \eta_{n+1}$, the hypothesis may be written as follows:

$$| \eta_n - \eta_{n+1} | \leq A. \tag{4.1}$$

Adding $| \eta_{n+1} |$ to both sides of (4.1), and recalling the triangle inequality, one obtains

$$| \eta_n | \leq A + | \eta_{n+1} |. \tag{4.2}$$

Since $| \eta_{n+1} | \leq pM_n$, (4.2) yields

$$| \eta_n | \leq A + pM_n. \tag{4.3}$$

Since the right-hand side of (4.3) is independent of the point at which η_n is evaluated, $| \eta_n |$ can be replaced by M_n, so that one obtains

$$M_n(1 - p) \leq A. \tag{4.4}$$

Dividing both sides of (4.4) by the positive quantity $1 - p$, we obtain the statement of the theorem.

THEOREM 4.2. If $\max | f_n - f_{n+1} | = A$, then $M_n = \max | \eta_n | \geq A/(1 + p)$.

PROOF. Given any $\epsilon > 0$, there can be found a value of s, say $s = s_0$, for which

$$| f_n(s_0) - f_{n+1}(s_0) | = | \eta_n(s_0) - \eta_{n+1}(s_0) | \geq A - \epsilon. \tag{4.5}$$

Now, since $| \eta_n(s_0) - \eta_{n+1}(s_0) | \leq | \eta_n(s_0) | + | \eta_{n+1}(s_0) |$, one obtains from (4.5) the further inequality

$$| \eta_n(s_0) | + | \eta_{n+1}(s_0) | \geq A - \epsilon. \tag{4.6}$$

But $| \eta_n(s_0) | \leq M_n$, $| \eta_{n+1}(s_0) | \leq M_{n+1} \leq pM_n$. Therefore,

[10] If any other choice of f_0 is made, the best upper bound that can be given for M_0 entirely on the basis of the principle of maximum and minimum will exceed 1.

[11] Here the additional difficulty is overlooked that at each stage of an actual computation it will almost certainly be necessary to approximate f_n by a simpler function in order to evaluate the succeeding function f_{n+1}, thus increasing the best bound that can be given for the maximum error.

[12] In Theorems 4.1 and 4.2 the term "max" stands for "least upper bound", as in Sec. 3.

$$M_n(1 + p) \geq A - \epsilon. \tag{4.7}$$

On account of the arbitrariness of the positive number ϵ, (4.7) implies that

$$M_n(1 + p) \geq A. \tag{4.8}$$

Dividing both sides of (4.8) by $1 + p$, we obtain the statement of Theorem 4.2.

THEOREM 4.3. If $f_n \leq f_{n+1}$ everywhere, then for $m > n$, $f_m \leq f_{m+1} \leq f$.

PROOF. The hypothesis may be written in the form

$$\eta_n - \eta_{n+1} \leq 0. \tag{4.9}$$

From (3.9) it is apparent that

$$\eta_{n+1} - \eta_{n+2} = \int (\eta_n - \eta_{n+1}) K_R \, ds'. \tag{4.10}$$

Since K_R is everywhere non-negative, it is clear that $\eta_{n+1} - \eta_{n+2} \leq 0$, and hence $f_{n+1} \leq f_{n+2}$. By repeating this argument, one obtains $f_m \leq f_{m+1}$ for $m \geq n$. Since $f_m \to f$, it follows that each estimate f_m lies below f.

This theorem obviously holds also with the inequalities reversed. Therefore, if two initial estimates $f_0^{(1)}$ and $f_0^{(2)}$ are chosen such that $f_0^{(1)} \leq f_1^{(1)}$ and $f_0^{(2)} \geq f_1^{(2)}$, the solution f of (2.6) is "squeezed" between the terms of the sequences $\{f_n^{(1)}\}$ and $\{f_n^{(2)}\}$.

***

5. Discussion of the inequality (3.3). In Sec. 3 a procedure was presented for the solution of the integral equation (2.6). The discussion was incomplete in one respect, for it was necessary to assume that the function $R(s)$ can be so chosen that the inequality (3.3) is satisfied. In Sec. 8 it will be shown, by an explicit computation, that (3.3) can be satisfied in the particular case of a domain bounded by any finite number of collinear slits. It appears, however, that in order to extend this result to more general types of domains it is necessary to impose certain restrictions upon the given domain and on the added arcs. Rather than attempting to state very general conditions, we shall here consider a rather restricted type of domain, but it will be apparent from the argument how one might proceed in the investigation of other domains.

It will be recalled from Sec. 2 that the integral $\int K_R(s, s') \, ds'$, extended over the added arcs, is simply the mean value over the circumference of the circle $C_R(P)$ (where P is the inner point of the added arcs corresponding to the parameter s) of the harmonic measure of the added arcs.[13] If the function $R(s)$ is continuous on each of the added arcs, as we shall assume, then the above integral is also a continuous function of s, and is less than unity, for the harmonic measure of the added arcs is never actually equal to unity (except possibly at the finite number of points where the added arcs meet the circumference). Therefore, in any particular case it is only necessary to show that the mean value is bounded away from unity in some neighborhood of each endpoint of the added arcs.

We shall now assume that the original (bounded) domain S and the added arcs CD and EF, as shown in Fig. 2, are such that the tangent to the added arc at each point P has no points in common with one of the two subdomains (S_2 in the figure) and meets

[13]The harmonic measure is taken, we recall, with respect to the subdomain in which the variable point on the circumference lies, not with respect to the original domain.

the boundary of the other subdomain (S_1) only at two points A and B (in addition to the point of tangency P). The points A and B will vary, of course, with the point P. Now, the mean value of the harmonic measure of the added arcs CD and EF will surely be bounded below unity as P varies along these arcs if the mean value on the semi-circle lying in the region[14] S_P bounded by the segment APB and the arc AGB is bounded

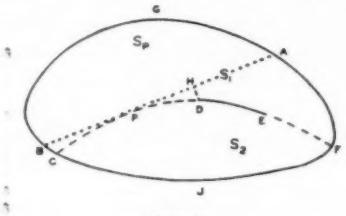

FIG. 2.

below unity, for the least upper bound of the mean value on the other semi-circle cannot exceed unity. Also, the harmonic measure of the added arcs at any point of the former semi-circle[15] is less than that of the curve $BCDEFA$, which, in turn, by the so-called "principle of extension" [3, p. 63], is less than that of the line segment AB with respect to the domain S_P ; the latter harmonic measure, furthermore, is less than that of the segment AB with respect to the entire half-plane (containing S_P) determined by extending AB to infinity in both directions. Now, we make the additional assumption that, as $P \to C$, R can be chosen $= c\overline{PB}$, where c is a sufficiently small positive constant. (This would not be possible, for example, if the arcs CD and CJF meet at C under zero angle.) Then, as is shown in Sec. 8, the mean value on the semi-circle of the harmonic measure of AB with respect to the half-plane will be bounded below unity.[16] Therefore, the inequality (3.3) holds as $P \to C$.

We now turn to the case when $P \to D$. Since the distance (in the usual point-set sense) between $C_R(P)$ and the arc EF is bounded away from zero, we may disregard the harmonic measure of this arc, and consider only the mean value over the semi-circle lying in S_P of the harmonic measure of the arc CD (with respect to S_1). We now assume that the arc CD is continuously curved in a neighborhood of D, so that the length of the segment DH of the perpendicular to AB from D satisfies, for some sufficiently large positive number σ and for P sufficiently close to D, the inequality

$$\overline{DH} \leq \sigma \overline{PH}^2. \tag{5.1}$$

We assume further that the point P and the entire arc DEF lie on opposite sides of the line determined by D and H. Then, by the aforementioned "principle of extension", the mean value over the semi-circle lying in S_P of the harmonic measure of the arc CD with respect to S_1 is less than the mean value of the harmonic measure of the broken line segment BHD with respect to the three-quarter-plane determined by extending

[14] The subscript P serves as a reminder that this domain depends upon the point P.

[15] For this semi-circle the harmonic measure will always be taken with respect to S_1 .

[16] The fact that the segment AB changes direction as P varies obviously causes no difficulty. The change in length of AB can be eliminated from our considerations by considering the harmonic measure of a (finite) segment, beginning at B, whose length exceeds the maximum length of AB. The considerations of Sec. 8 are then immediately applicable.

HB and HD beyond B and D respectively to infinity. See Fig. 3. This mean value, in turn, is majorized by the mean value of the harmonic measure (still with respect to the three-quarter-plane) of the entire half-line $HB\infty$ plus that of HJ, where $\overline{HJ}$ is equal

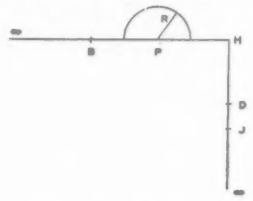

FIG. 3.

to the right-hand side of (5.1).[17] (As pointed out previously, we may assume $\overline{PH}$ so small that (5.1) holds.) We now assume that R is chosen as follows:

$$R = \gamma\,\overline{PH}, \tag{5.2}$$

where γ is some positive constant <1. Then, from similarity considerations, it is apparent that the mean value of the harmonic measure of $HB\infty$ remains constant (and therefore certainly <1) as P approaches D. It also follows from similarity considerations, together with (5.1) and (5.2), that the contribution from HJ, and therefore that from HD, diminishes to zero as P approaches D. Therefore the mean value of the harmonic measure of DHB is certainly bounded away from unity. From the chain of inequalities which we have established it follows a fortiori that $\int K_R(s, s')\,ds'$ is bounded away from unity as P approaches D.

Finally, we note that the discussions of the cases $P \to E$ and $P \to F$ are entirely parallel to those for $P \to D$ and $P \to C$ respectively.

While the discussion presented here deals with a rather specialized type of domain, it seems likely that attempts to prove that the inequality (3.3) can be satisfied for a given domain must in almost all cases proceed along the lines used here—i.e., of repeated use of the "principle of extension"; for in only very few cases (of which that considered in Secs. 6 and 8 is one) is it possible to evaluate the integral $\int K_R(s, s')\,ds'$ in closed form.

***

6. Specific application to "slit case". In this section the method described in the first part of this paper will be applied to a rather special type of domain which is, however, of considerable interest.

Let the domain S for which the Dirichlet problem is to be solved consist of the entire (x, y)-plane except for a finite number, n, of collinear slits, which constitute the boundary of S. Without loss of generality (since the Laplace equation is invariant under rotation and translation) these slits may be considered to lie on the x-axis. Also it may be assumed that the "point at infinity" is an inner[18] boundary point of S, for this can always be realized by performing an inversion in a circle with center at any inner point of any one of the slits. (The two-dimensional Laplace equation is invariant under inversion

[17] The orientation of the boundary of the new domain will vary with P, but this does not cause any difficulty.

[18] That is, the slits extend to infinity both on the right and on the left.

as well as rotation and translation.) The portion of the x-axis consisting of the slits will henceforth be designated as C and the remainder as D (which consists of the n intervals $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)$). There is now defined on C a function[19] $h(x)$, which is assumed to be bounded on each slit and continuous except perhaps at a finite number of points. The theorem referred to in Sec. 2 from the Nevanlinna treatise is directly applicable, so that we are assured of the existence of a unique function $u(P)$, $P = (x, y)$, which is harmonic and bounded in S and approaches $h(x)$ whenever P approaches a point of C at which $h(x)$ is continuous.

Now, it is clear that the problem would be essentially solved if the values of $u(P)$ could be determined along each of the intervals $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)$ constituting D, for then the values of $u(P)$ could be determined at all points not on the x-axis by the Poisson integral formula:

$$u(P) = \frac{|y|}{\pi} \int_{-\infty}^{\infty} \frac{u(\xi, 0)\, d\xi}{(\xi - x)^2 + y^2}, \qquad (y \neq 0). \tag{6.1}$$

Letting $u(x, 0)$ on D be designated as the function $f(x)$, we may rewrite (6.1) in the form

$$u(P) = \frac{|y|}{\pi} \int_C \frac{h(\xi)\, d\xi}{(\xi - x)^2 + y^2} + \frac{|y|}{\pi} \int_D \frac{f(\xi)\, d\xi}{(\xi - x)^2 + y^2}, \qquad (y \neq 0). \tag{6.2}$$

Let $(x, 0)$ be any point, temporarily fixed, interior to D, say inside the interval (a_k, b_k) and let R be any positive number subject only to the condition

$$R \leq \min (x - a_k, b_k - x); \tag{6.3}$$

then the mean-value property for $u(P)$ at the point $(x, 0)$ may be expressed (on account of the evenness of $u(P)$ in y, which is apparent from (6.1)) as follows:

$$f(x) = \frac{1}{\pi} \int_0^{\pi} u(x + R \cos \theta, R \sin \theta)\, d\theta. \tag{6.4}$$

Now, for $0 < \theta < \pi$, equation (6.2) may be written in the form

$$\begin{aligned} u(x + R \cos \theta, R \sin \theta) = {} & \frac{R \sin \theta}{\pi} \int_C \frac{h(\xi)\, d\xi}{(\xi - x - R \cos \theta)^2 + R^2 \sin^2 \theta} \\ & + \frac{R \sin \theta}{\pi} \int_D \frac{f(\xi)\, d\xi}{(\xi - x - R \cos \theta)^2 + R^2 \sin^2 \theta}. \end{aligned} \tag{6.5}$$

We now proceed as explained in Sec. 2. The integrand of (6.4) is replaced by the right-hand side of (6.5), the order of integration is inverted, and the inner integration is performed. In this manner, we obtain

[19] More generally, different functions could be prescribed on the upper and lower edges of C. (In the discussion of the more general case in Sec. 2 this possibility was considered.) If different boundary values $h^+(x)$ and $h^-(x)$ were prescribed on the upper and lower edges of C respectively, the problem could easily be reduced to one of the type considered here by dividing it into two problems, one with boundary values $\frac{1}{2}(h^+(x) + h^-(x))$ on both edges of C, and the other with boundary values $\frac{1}{2}(h^+(x) - h^-(x))$ and $\frac{1}{2}(h^-(x)' - h^+(x))$ on the upper and lower edges respectively. For the second problem the solution must obviously vanish on D (by symmetry considerations), so that by using the Poisson formula for the upper and lower half-planes in turn the second problem would be solved, leaving only a problem of the kind considered here.

$$f(x) = \int_C h(\xi)K(x, R, \xi)\, d\xi + \int_D f(\xi)K(x, R, \xi)\, d\xi, \tag{6.6}$$

where

$$K(x, R, \xi) = (\pi^2(\xi - x))^{-1} \log \left| \frac{\xi - x + R}{\xi - x - R} \right|. \tag{6.7}$$

(When $\xi = x$, $K(x, R, \xi)$ is to be taken equal to $2/\pi^2 R$.)

Now, for each value of x interior to D let a positive R be chosen, subject only to the restriction (6.3); i.e., a function $R(x)$ is chosen. Then the first term on the right-hand side of (6.6) becomes a definite function of x, defined in the interior of D. This function will be designated $g_R(x)$, the subscript R indicating, as in Sec. 2, that this function depends on the choice made for the function $R(x)$. For ξ interior to D, the right-hand side of (6.7) becomes a definite function of ξ and x, with x also assuming values in the interior of D. This function will be designated $K_R(x, \xi)$. Thus, there is obtained for $f(x)$ an integral equation (or, in a sense, a family of integral equations, one for each possible choice of the function $R(x)$):

$$f(x) = g_R(x) + \int_D K_R(x, \xi)f(\xi)\, d\xi. \tag{6.8}$$

In Sec. 8 we shall first show that the function $R(x)$ can be so chosen that condition (3.3) is satisfied, so that the method prescribed in Sec. 3 for the solution of (6.8) (or, more generally (2.6)) is applicable, and then shall illustrate our method by means of two numerical examples. Before this, however, we shall present in Sec. 7 a brief discussion of the problem of choosing a suitable approximation with which to begin the procedure described in Sec. 3.

7. Choice of an initial approximation. Although the sequence of functions $\{f_n(s)\}$ obtained by the procedure described in Sec. 3 converges to $f(s)$, the solution of the integral equation (2.6), regardless of whether $f_0(s)$ is a good or poor approximation to $f(s)$, it is essential from the practical viewpoint to be able to select a suitable initial approximation, for with a poor choice of $f_0(s)$ the number of succeeding functions $f_1(s), f_2(s), \cdots$ which would be necessary to provide a sufficiently accurate approximation to $f(s)$ might be so large as to make our procedure useless for computational purposes. For simplicity we shall restrict ourselves to the consideration of domains of the type considered in Sec. 6, and, correspondingly, to the problem of solving the integral Eq. (6.8), whose kernel is defined by (6.7); it will be apparent, however, that the basic idea which we shall employ is also applicable in the general case.

It is obvious that Eq. (6.8) may be expressed in variational terms as the problem of finding a function $f(x)$ for which

$$\int_D \left(f(x) - g_R(x) - \int_D f(\xi)K_R(x, \xi)\, d\xi \right)^2 dx = \text{minimum}, \tag{7.1}$$

where the class of admissible functions may be taken as that of all functions which are bounded and continuous in the interior of each interval of D. Now, following the method of Rayleigh-Ritz [1, pp. 149–151], we would replace $f(x)$ in (7.1) by a function $F(x, \alpha, \beta, \gamma, \cdots)$ of x and a number of parameters $\alpha, \beta, \gamma, \cdots$ (and correspondingly replace $f(\xi)$ by $F(\xi, \alpha, \beta, \gamma, \cdots)$), and then choose the parameters so as to minimize

the left-hand side of (7.1). The function obtained with this choice of parameters is then the best approximate solution to (7.1) among the class of functions considered, and may therefore reasonably be expected to serve as a good initial approximation for the procedure described in Sec. 3. Actually, from Theorems 4.1 and 4.2 it is seen that it would be preferable to select the parameters so as to minimize

$$\max_{x \in D} \left| F(x, \alpha, \beta, \gamma, \cdots) - g_R(x) - \int_D F(\xi, \alpha, \beta, \gamma, \cdots) K_R(x, \xi)\, d\xi \right|, \tag{7.2}$$

but it is evident that this is in general a hopelessly difficult task. Whether or not the initial approximation obtained by minimizing the left-hand side of (7.1) is actually a good one can *then* be seen by evaluating the expression (7.2).

In particular, we may take $F(x, \alpha, \beta, \gamma, \cdots)$ to be a polynomial in x whose coefficients are the parameters α, β, γ, $\cdots$ (cf. Example 1, Sec. 8) or, more generally, $F(x, \alpha, \beta, \gamma, \cdots)$ may consist of different polynomials in the different intervals of D (cf. Example 2, Sec. 8). As may be seen by performing the substitution

$$\xi = x + Ru, \tag{7.3}$$

the integral

$$\int_D F(\xi, \alpha, \beta, \gamma, \cdots) K_R(x, \xi)\, d\xi \tag{7.4}$$

can be expressed as a sum of integrals of elementary type plus a single non-elementary integral, namely

$$\int u^{-1} \log \left| \frac{1+u}{1-u} \right| du, \tag{7.5}$$

which can be easily evaluated by series methods. However, even when $F(x, \alpha, \beta, \gamma, \cdots)$ is of the simple form considered here, the minimizing of the left-hand side of (7.1) is not a simple task, for the integration with respect to x involves non-elementary functions. Nevertheless, the best values of the parameters α, β, γ, $\cdots$ may be determined at least approximately by replacing the integration with respect to x by a summation, say by using Simpson's rule. This is the method which was employed in Sec. 8. In Example 1, the interval $-1 \leq x \leq 1$ was divided into 20 equal subintervals, and in Example 2 each of the intervals $-4 \leq x \leq -1$ and $1 \leq x \leq 4$ was divided into 12 equal sub-intervals.

The fact that very satisfactory initial approximations were obtained in Sec. 8 by the method which we have described here suggests that the variational approach is a suitable means for obtaining an initial approximation, and also that the approximate method described in the above paragraph for obtaining the best values of the parameters yields values not differing appreciably from the correct ones.

8. Illustrative examples for domains bounded by collinear slits. In proving the convergence of the approximation procedure in Sec. 3 we had to assume that the inequality (3.3) is satisfied. In this section we shall first of all show that, for the domains considered in Sec. 6, it is indeed possible to assure, by a suitable choice of the function $R(x)$, that (3.3) holds.

The function $R(x)$ will be chosen as large as possible. Therefore,

$$R(x) = \min (x - a_k, b_k - x), \tag{8.1}$$

where x lies in the interior of the interval (a_k , b_k). In order to show that with this choice of $R(x)$ a constant p exists for which (3.3) holds, we consider the function

$$p(x) = \int_D K_R(x, \xi)\, d\xi = \frac{1}{\pi^2} \int_D (\xi - x)^{-1} \log \left| \frac{\xi - x + R}{\xi - x - R} \right| d\xi, \tag{8.2}$$

defined for all values of x interior to D. If x lies in the left half of the k'th interval, (a_k , b_k), so that $R = x - a_k$, we may write (8.2) in the form:

$$p(x) = \frac{1}{\pi^2} \int_D (\xi - x)^{-1} \log \left| \frac{\xi - a_k}{\xi - 2x + a_k} \right| d\xi. \tag{8.3}$$

Since the integrand is positive everywhere, the right-hand side of (8.3) is increased if the integration is extended over a point set containing D. Extending the integration over all values of ξ except for the segment of C immediately to the left of the k'th interval[20] of D, we therefore obtain the inequality

$$\begin{aligned} p(x) < \frac{1}{\pi^2} \int_{-\infty}^{\infty} (\xi - x)^{-1} \log \left| \frac{\xi - a_k}{\xi - 2x + a_k} \right| d\xi \\ - \frac{1}{\pi^2} \int_{b_{k-1}}^{a_k} (\xi - x)^{-1} \log \left| \frac{\xi - a_k}{\xi - 2x + a_k} \right| d\xi. \end{aligned} \tag{8.4}$$

Performing in (8.4) the substitution

$$\xi = x + Ru = x + (x - a_k)u, \tag{8.5}$$

we obtain

$$p(x) < \frac{1}{\pi^2} \int_{-\infty}^{\infty} u^{-1} \log \left| \frac{1 + u}{1 - u} \right| du - \frac{1}{\pi^2} \int_{(b_{k-1} - x)/(x - a_k)}^{-1} u^{-1} \log \left| \frac{1 + u}{1 - u} \right| du. \tag{8.6}$$

Now, it is known[21] that

$$\frac{1}{\pi^2} \int_{-\infty}^{\infty} u^{-1} \log \left| \frac{1 + u}{1 - u} \right| du = 1. \tag{8.7}$$

Also, since the integrand of the second term on the right-hand side of (8.6) is always positive, the entire right-hand side is largest when the lower limit of integration is as large (i.e., as small in absolute value) as possible. This clearly occurs when x is as large as possible, i.e., when $x = 1/2(a_k + b_k)$, so that, for $a_k < x \leq 1/2(a_k + b_k)$, the following inequality holds:

$$p(x) \leq 1 - \frac{1}{\pi^2} \int_{[2b_{k-1} - (a_k + b_k)]/(b_k - a_k)}^{-1} u^{-1} \log \left| \frac{1 + u}{1 - u} \right| du. \tag{8.8}$$

---

[20] If $k = 1$, the interval $(-\infty, a_1)$ is to be excluded from the integration. In (8.4), b_0 is therefore to be taken as $-\infty$.

[21] This can be seen without performing any computations, in the following way. By taking account of the substitution (8.5), one sees that the left-hand side of (8.7) is equal to the first term on the right-hand side of (8.4). The latter quantity, in turn, is, by the considerations of Sec. 6, equal to the mean value over a certain semi-circle of the harmonic measure of the entire x-axis with respect to the half-plane $y = 0$. Since the harmonic measure of the entire boundary of any domain is identically equal to one at all points of the domain, (8.7) follows.

Similarly, it is possible to obtain for $p(x)$ an upper bound less than one in the right half of each interval of D. If p is taken equal to the largest of the $2n$ upper bounds so obtained, the inequality (3.3) will be satisfied.

A similar proof can be carried out for various other choices of $R(x)$; e.g., one may choose $R(x) = c \min (x - a_k, b_k - x)$, where c is any positive constant <1. However, the value of p obtained with any $c < 1$ would be closer to one than with $c = 1$, so that the choice $c = 1$ appears best. In the numerical examples to be presented here $R(x)$ is always chosen with $c = 1$, i.e., according to (8.1). In Example 1, we have $p = .75$; in Example 2, $p = .8387$.

EXAMPLE 1. The domain S consists of the entire (x, y)-plane cut by the slits $(-\infty, -1)$ and $(1, \infty)$ of the x-axis, on which the boundary values $h(x) = -1$ and $h(x) = 1$ respectively are prescribed.

The exact solution is easily found to be given, on the interval $(-1, 1)$ of the x-axis, by

$$u(x, 0) = f(x) = \frac{2}{\pi} \arcsin x; \tag{8.9}$$

however, in order to illustrate our procedure, we have taken three different initial approximations, which we designate $f_0^{(1)}(x)$, $f_0^{(2)}(x)$, $f_0^{(3)}(x)$, and have determined the corresponding succeeding approximations, which we designate $f_1^{(1)}(x)$, $f_1^{(2)}(x)$, $f_1^{(3)}(x)$, respectively. The results of these computations are presented in Table 1. (Only positive values of x are listed, since all functions appearing are odd.)

TABLE 1

x	$f_0(x)$	$f_1(x)$	$f_0(x)$	$f_1(x)$	$f_0(x)$	$f_1(x)$	$f(x)$
0	0	0	0	0	0	0	0
.1	.1	.0589	.0600	.0624	.0600	.0652	.0638
.2	.2	.1268	.1215	.1281	.1215	.1317	.1282
.3	.3	.2042	.1860	.1935	.1860	.1979	.1940
.4	.4	.2894	.2550	.2613	.2550	.2665	.2620
.5	.5	.3819	.3300	.3317	.3300	.3375	.3333
.6	.6	.4820	.4126	.4069	.4126	.4113	.4097
.7	.7	.5920	.5041	.4907	.5041	.4977	.4936
.8	.8	.7072	.6062	.5885	.6062	.5976	.5903
.9	.9	.8388	.7203	.7098	.7203	.7250	.7129
1.0	1.	1.	.8479	.8859	1.	1.	1.

The first initial approximation to be chosen was:

$$f_0^{(1)}(x) = x. \tag{8.10}$$

This function, although joining continuously with the prescribed boundary values, is clearly a poor choice; nevertheless, as may be seen from Table 1, the succeeding approximation, $f_1^{(1)}(x)$, shows a very marked tendency toward the exact solution given by (8.9).

The function $f_0^{(2)}(x)$ was chosen to be the "best" third-degree polynomial (in the sense of Sec. 7). We thus obtained:

$$f_0^{(2)}(x) = .5974x + .2505x^3. \tag{8.11}$$

This function is seen to be a rather good approximation to the exact solution except near the end-points of the range, for it fails to join continuously with the prescribed boundary-values. The succeeding approximation $f_1^{(2)}(x)$ is seen to be still closer to $f(x)$ for most values of x, and the improvement near the ends is worthy of note.

The third choice of an initial approximation was made as follows: The function (8.11) was retained in the sub-interval $-.9 \leq x \leq .9$, but for $-1 \leq x \leq -.9$ and for $.9 \leq x \leq 1$ was replaced by second-degree polynomials joining smoothly with (8.11) at $x = \pm.9$ and continuously with the prescribed boundary values at $x = \pm 1$. In this way the following choice was obtained for $f_0^{(3)}(x)$:

$$\begin{aligned}
&\text{(a)} \quad f_0^{(3)}(x) = -12.5230 - 27.4344x - 15.9114x^2, && -1 \leq x \leq -.9; \\
&\text{(b)} \quad f_0^{(3)}(x) = f_0^{(2)}(x) = .5974x + .2505x^3, && -.9 \leq x \leq .9; \\
&\text{(c)} \quad f_0^{(3)}(x) = 12.5230 - 27.4344x + 15.9114x^2, && .9 \leq x \leq 1.
\end{aligned} \tag{8.12}$$

The corresponding approximation $f_1^{(3)}(x)$ is seen to provide a remarkably good approximation to $f(x)$ throughout the range $(-1, 1)$.

EXAMPLE 2. The domain S consists of the entire (x, y)-plane cut by the slits $(-\infty, -4)$, $(-1, 1)$, and $(4, \infty)$ of the x-axis, on which the boundary values $h(x) = 0, 1, 0$ respectively are prescribed.[22]

The exact solution to this problem can be expressed in terms of elliptic integrals. In particular, on the interval $(1, 4)$ of the x-axis, the solution is given by

$$u(x, 0) = f(x) = \frac{1}{2} - \frac{\int_0^w (1 - t^2/81)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\, dt}{2\int_0^1 (1 - t^2/81)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\, dt}, \tag{8.13}$$

where

$$w = \frac{3(x - 2)}{x + 1}. \tag{8.14}$$

In the notation commonly used for elliptic integrals, (8.13) may be written in the form

$$f(x) = \frac{1}{2} - \frac{K(\arcsin w,\ 1/9)}{2K(1/9)}. \tag{8.15}$$

(Since the solution is obviously even in x, it suffices to determine $f(x)$ in the interval $(1, 4)$.)

As an initial approximation we first tried

$$f_0^{(1)}(x) = \frac{1}{3}(4 - |\, x\,|), \tag{8.16}$$

which is seen to join continuously with the prescribed boundary values. (As in Example 1, a superscript is attached to the approximation function in order to distinguish it from another one which will be used later.) As in the case of the first initial approximation considered in Example 1, this is a rather poor choice; nevertheless, as may be seen from

[22] That is, the problem is that of determining the harmonic measure of the finite slit.

Table 2, the succeeding approximation $f_1^{(1)}(x)$ shows a strong tendency toward the exact solution $f(x)$, whose values are also given in Table 2.

TABLE 2

x	$f_0(x)$	$f_1(x)$	$f_0(x)$	$f_1(x)$	$f(x)$
$1+0$	1.	1.	.8985	.9239	1.
1.25	.9167	.8549	.7629	.7471	.7430
1.5	.8333	.7499	.6537	.6419	.6402
1.75	.7500	.6535	.5664	.5648	.5639
2.0	.6667	.5687	.4966	.5014	.5000
2.25	.5833	.4956	.4397	.4443	.4438
2.5	.5000	.4366	.3913	.3925	.3921
2.75	.4167	.3895	.3470	.3442	.3434
3.0	.3333	.3461	.3023	.3002	.2956
3.25	.2500	.2819	.2527	.2468	.2472
3.50	.1667	.2062	.1937	.1953	.1954
3.75	.0833	.1168	.1209	.1309	.1342
4–0	0	0	.0300	.0225	0

As in Example 1, we next obtained another approximation by the method described in Sec. 7. The approximation was taken to be of the form

$$\begin{aligned} f_0^{(2)}(x) &= a + bx + cx^2 + dx^3, \qquad 1 \leq x \leq 4; \\ f_0^{(2)}(x) &= a - bx + cx^2 - dx^3, \qquad -4 \leq x \leq -1, \end{aligned} \tag{8.17}$$

the coefficients a, b, c, d being unspecified to begin with. (From (8.17) it is apparent that $f_0^{(2)}(x)$ is even, so that as before it suffices to determine $f_1^{(2)}(x)$ for $1 \leq x \leq 4$.) In this manner the following values were found for the coefficients:

$$a = 1.7936, \qquad b = -1.2369, \qquad c = .3894, \qquad d = -.0476. \tag{8.18}$$

With this choice of the coefficients, the values of $f_0^{(2)}(x)$ and $f_1^{(2)}(x)$ listed in Table 2 were obtained. As in the corresponding part of Example 1, these functions are seen to be rather good approximations to the exact solution, except near the end-points. It is almost certain that if $f_0^{(2)}(x)$ were modified near the end-points so as to join continuously with the prescribed boundary-values, as was done in Example 1, both the initial and succeeding approximations would constitute considerable improvements over the ones which we have obtained.

BIBLIOGRAPHY

1. Courant-Hilbert, *Methoden der mathematischen Physik*, Vol. I, 2nd ed., 1931.
2. Courant-Hilbert, *Methoden der mathematischen Physik*, Vol. II, 1937.
3. R. Nevanlinna, *Eindeutige analytische Funktionen*, 1936.
4. R. Nevanlinna, Crelles Journal 180, 121-128 (1938).
5. S. Saks, *Theory of the integral*, 1937.

—NOTES—

ON A FAMILY OF ROTATIONAL GAS FLOWS*

By R. C. PRIM (*Naval Ordnance Laboratory*)

Steady plane flows of an ideal gas (i.e., a thermodynamically perfect gas without viscosity or thermal conductivity) in the absence of body forces may be divided into three distinct classes:

(a) flows in which the velocity field **v** is irrotational,

(b) flows in which the **v** field is rotational, but is derivable from an irrotational flow by the substitution principle established by Munk and Prim[1], and

(c) all other flows.

For the study of rotational gas flows, the actual velocity field **v** does not play so basic a role as does the reduced velocity field **w** defined by[1,2]

$$\mathbf{w} \equiv \mathbf{v}/a,$$

where a is the local value of the maximum or "ultimate" velocity magnitude attainable by expansion to zero pressure. A necessary condition for irrotationality of **v** is that a be constant throughout the flow. Therefore the flows in Class (a) are characterized by irrotationality of both the **w** field and the **v** field. The flows in Class (b) have rotational **v** fields, but share their **w** field (and hence their geometrical properties) with flows of Class (a). The flows of Class (c) are truly rotational, i.e. they have rotational **w** fields as well as rotational **v** fields and will be, in general, geometrically distinct from the flows of Class (a).

Of the truly rotational flows of Class (c), those having a streamline pattern consisting of concentric circles or of parallel straight lines constitute a relatively degenerate subclass. Any value of velocity magnitude v and of reduced velocity magnitude w may be assigned to each individual streamline, and the resulting flow will satisfy all the governing equations. Except for this degenerate case, no formally simple solutions for this class of flows have been known heretofore.

This paper presents a formally simple infinite family of truly rotational flow solutions. This family includes the familiar irrotational Prandtl-Meyer corner flow as a special case.

The new family of solutions. Necessary and sufficient conditions that a vector field must satisfy in order that it may serve as the reduced velocity field of a steady flow of an ideal gas in the absence of body forces are the following set of equations[2]

$$\operatorname{div}[(1 - w^2)^{(\lambda-1)/2}\mathbf{w}] = 0, \tag{1}$$

$$\operatorname{curl}\left[\frac{\mathbf{w} \times \operatorname{curl} \mathbf{w}}{1 - w^2}\right] = 0. \tag{2}$$

*Received Nov. 24, 1947.

[1] M. Munk and R. Prim, *On the multiplicity of steady gas flows having the same streamline pattern*, Proc. Nat. Acad. Sci., **33**, 137-141 (1947).

[2] M. Munk and R. Prim, *On the canonical form of the equations of steady motion of a perfect gas*, Naval Ordnance Laboratory Memorandum 9169 (1947).

($\lambda \equiv (\gamma + 1)/(\gamma - 1)$, where γ denotes the adiabatic exponent; for air, $\lambda = 6$.)

We shall refer these equations to a cylindrical coordinate system r, θ, z and consider flows of the following type:

$$w_r = u(\theta), \qquad w_\theta = v(\theta), \qquad w_z = 0.$$

Thus specialized, Eqs. (1) and (2) yield the following pair of ordinary differential equations restricting the functions $u(\theta)$ and $v(\theta)$

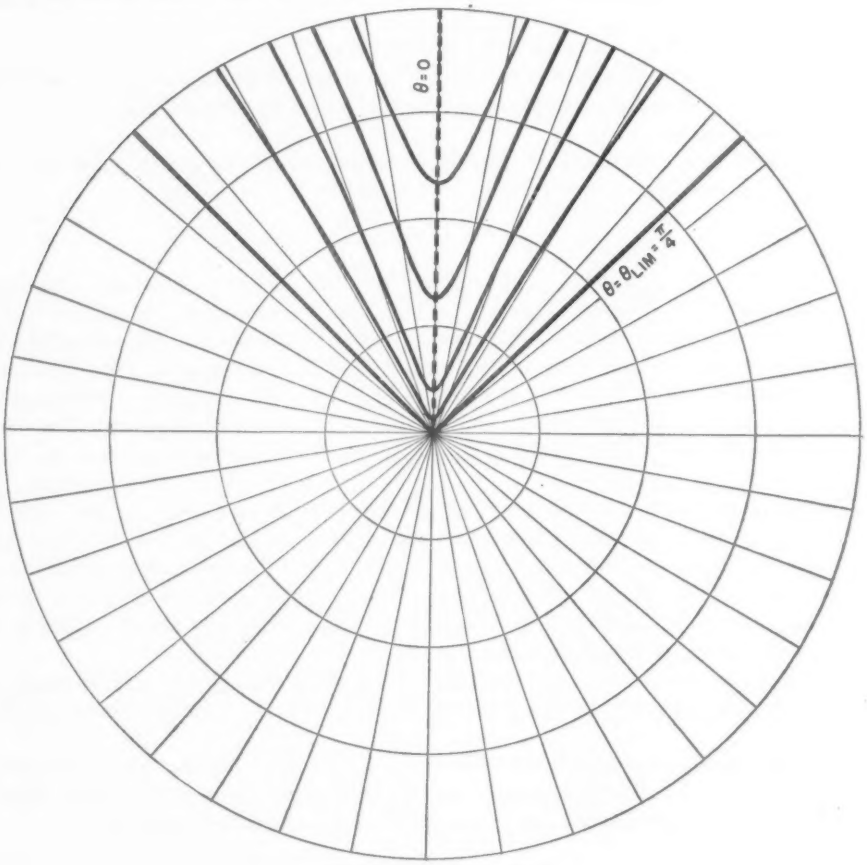

FIG. 1.

$$(u + v')(1 - u^2 - v^2) - (\lambda - 1)v(uu' + vv') = 0, \tag{3}$$

$$v(v - u') - B(1 - u^2 - v^2) = 0, \tag{4}$$

where B is an arbitrary constant parameter. (These equations appear in more general form in a paper by Nemenyi and Prim.[3])

[3]P. Nemenyi and R. Prim, *Some patterns of vorticose flow of a perfect gas*, Naval Ordnance Laboratory Memorandum 9219 (1947).

The general solution of the set of equations (3) and (4) is formally forbidding and yields the functions $u(\theta)$ and $v(\theta)$ only implicitly. However, for the case $B = 0$ (for which the w field is irrotational) the solution is easy and yields the following two types of solutions:

$$v = \lambda^{-1/2} \cos [(\theta - \theta_0)/\lambda^{1/2}], \qquad u = \sin [(\theta - \theta_0)/\lambda^{1/2}], \tag{5}$$

and

$$v = A \cos (\theta - \theta_0), \qquad u = A \sin (\theta - \theta_0), \tag{6}$$

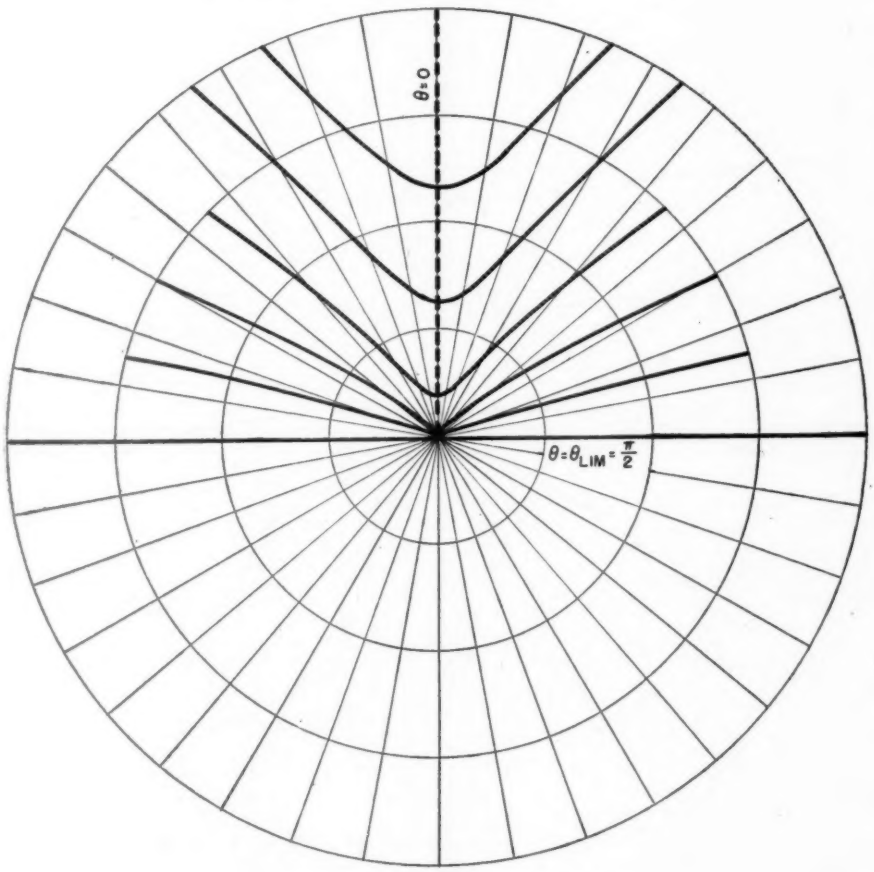

FIG. 2.

where θ_0 and A are arbitrary constants (except that $A^2 < 1$ physically). The solution (5) is the well-known Prandtl-Meyer irrotational corner flow, while solution (6) is the trivial case of uniform parallel flow with $w = A$.

Guided by the form of (5) and (6) we investigate the possibility of special solutions of the form

$$v = C \cos m(\theta - \theta_0), \qquad u = D \sin m(\theta - \theta_0), \tag{7}$$

when B is not required to be zero. Substitution of (7) into (3) and (4) yields

$$\cos^2 m(\theta - \theta_0) = - \frac{(D - Cm)(1 - D^2)}{(D^2 - C^2)(D - \lambda Cm)} \tag{8}$$

and

$$\cos^2 m(\theta - \theta_0) = - \frac{B(1 - D^2)}{B(D^2 - C^2) - C(C - Dm)}. \tag{9}$$

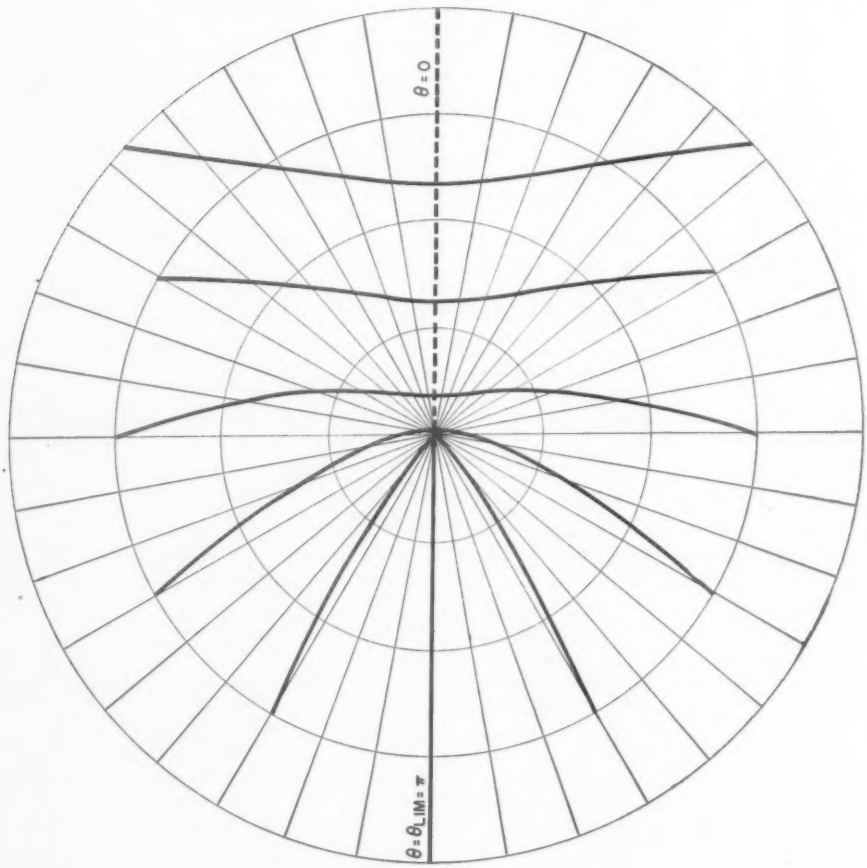

FIG. 3.

In order that (8) and (9) be valid for some finite range of $\theta - \theta_0$ it is necessary that the numerator and denominator of each right-hand member be identically zero. In addition to the trivial condition $C^2 = D^2 = m^2 = 1$, the following relationships between B, C, D and m satisfy these conditions

$$B = \frac{\lambda C^2 - 1}{\lambda(1 - C^2)}, \qquad D = \pm 1, \qquad m = \frac{D}{\lambda C}. \tag{10}$$

Replacing C by v_0 and setting $\theta_0 = 0$ (this involves no real loss of generality), we have then the solution

$$v = v_0 \cos(\theta/\lambda v_0), \qquad u = \sin(\theta/\lambda v_0) \tag{11}$$

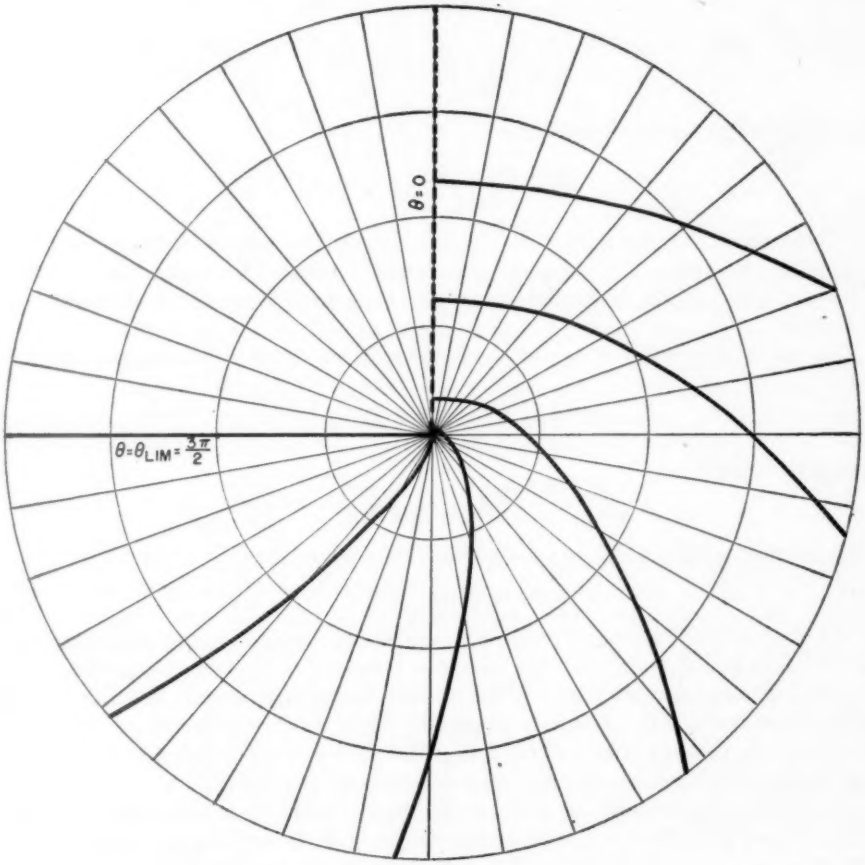

FIG. 4.

where v_0 is an arbitrary parameter (limited physically to the range $0 < v_0^2 < 1$).

This family of solutions is as simple formally as the Prandtl-Meyer solution (5) to which it reduces for the special case $v_0^2 = 1/\lambda$. For all other values of v_0 the solutions are truly rotational flows of Class (c).

Some properties of the solutions. The *equation of the streamlines* of these flows is readily obtained by integration of the defining equation

$$\frac{dr}{rd\theta} = \frac{u}{v} = \frac{1}{v_0} \tan(\theta/\lambda v_0)$$

giving

$$r = r_0 \left[\sec\left(\theta/\lambda v_0\right)\right]^{\lambda} \tag{12}$$

The *rotation* of the reduced velocity field is given by

$$\Omega \equiv \frac{1}{2} \mid \nabla \times \mathbf{w} \mid = \frac{v - u'}{2r} \tag{13}$$

$$\Omega = \frac{\lambda v_0^2 - 1}{2\lambda v_0} \frac{\cos\left(\theta/\lambda v_0\right)}{r}$$

or, along a particular streamline,

$$\Omega = \frac{\lambda v_0^2 - 1}{2\lambda v_0 r_0} \left[\cos\left(\theta/\lambda v_0\right)\right]^{\lambda+1}. \tag{14}$$

It will be noted that for $v_0 > 0$ the rotation is positive, zero, or negative near $\theta = 0$ depending on whether v_0^2 is respectively greater than, equal to, or less than $1/\lambda$, where $1/\lambda$ is the square of the "critical velocity" at which the flow velocity and sound velocity are equal.

The *pressure distribution* can be computed from the $\mathbf{w}$ field using the equation[2]

$$\operatorname{grad} \ln p = -(\lambda + 1) \frac{(\mathbf{w} \cdot \operatorname{grad})\mathbf{w}}{1 - w^2}$$

the result being

$$p = k\left\{r^{(\lambda v_0^2 - 1)/\lambda(1 - v_0^2)} \left[\cos\left(\theta/\lambda v_0\right)\right]^{(\lambda-1)v_0^2/(1-v_0^2)}\right\}^{\lambda+1} \tag{15}$$

from which is readily obtained the *equation of the isobars*

$$r = L\left[\sec\left(\theta/\lambda v_0\right)\right]^{\lambda(\lambda-1)v_0^2/(\lambda v_0^2 - 1)} \tag{16}$$

where K and L are arbitrary positive constants. It should be noted that only for the irrotational case ($v_0^2 = 1/\lambda$) are the lines, $\theta =$ constant, lines of constant pressure. From (15) it is also seen that the nature of the pressure distribution is quite different depending on whether the initial velocity is supersonic ($v_0^2 > 1/\lambda$) or subsonic ($v_0^2 < 1/\lambda$). For $v_0^2 > 1/\lambda$, the pressure is zero at the origin and increases with increasing r; for $v_0^2 < 1/\lambda$ the pressure is infinite at the origin and decreases with increasing r.

A familiar characteristic of the Prandtl-Meyer solution is the existence of a *limiting angle* ($\theta_{\text{lim}} = \pi\lambda^{1/2}/2$) at which the flow velocity is of ultimate magnitude and radial in direction ($u = 1$, $v = 0$). From (11) it is seen that for each member of our family of solutions there exists such a limiting angle, its magnitude depending on the choice of v_0:

$$\theta_{\text{lim}} = v_0 \lambda \pi/2. \tag{17}$$

By varying the choice of v_0 from zero to one, we may vary the limiting angle continuously from zero to $\lambda\pi/2$, (3π for air).

Another familiar characteristic of the Prandtl-Meyer solution is the fact that the tangential velocity component v is always equal to the local value of the velocity of sound c, i.e., $v/c = \pm 1$. We shall find that our entire family of solutions (11) has a similar, but more general, property which reduces to $v/c = \pm 1$ when $v_0^2 = 1/\lambda$. The square of the velocity of sound (referred to the ultimate velocity, of course) is given by

$$c^2 = \frac{1 - w^2}{\lambda - 1} \tag{18}$$

Making use of (11) we then have

$$\frac{v^2}{c^2} = v_0^2 \cos^2 \frac{\theta}{\lambda v_0} \left[\frac{1 - v_0^2}{\lambda - 1} \cos^2 \frac{\theta}{\lambda v_0} \right]^{-1} = \frac{(\lambda - 1)v_0^2}{1 - v_0^2}$$

or, since for steady flows of ideal gas in absence of body forces, the Mach number M is related to w through

$$M^2 = \frac{(\lambda - 1)w^2}{1 - w^2} \tag{19}$$

$$\frac{v}{c} = \pm M_0 , \tag{20}$$

where M_0 is the Mach number at $\theta = 0$. (For the Prandtl-Meyer solution, $M_0 = 1$).

It is noteworthy that only for the irrotational case ($v_0^2 = 1/\lambda$, $M_0 = 1$) are the radial lines characteristic lines of the differential equations. Therefore, the useful "patching" properties of the Prandtl-Meyer solution are not shared by the other members of our family of solutions. However, sufficiently restricted regimes of our flows could be obtained by passage of an initially irrotational flow through a curved shock front.

The *Mach number* at any point in the flow is, from (11) and (19) given by

$$M = \left[\frac{(\lambda - 1)[1 - (1 - v_0^2) \cos^2 (\theta/\lambda v_0)]}{(1 - v_0^2) \cos^2 (\theta/\lambda v_0)} \right]^{1/2} \tag{21}$$

By virtue of the substitution principle[1] each member of the family of solutions (11) in terms of the reduced velocity field **w** represents a vast variety of *actual velocity fields* **v**. Denoting the tangential and radial components of **v** respectively by v^* and u^*, and using r_0 of (12) to parametrize the streamlines, we have

$$\begin{aligned} v^* &= a(r_0) v_0 \cos (\theta/\lambda v_0) \\ u^* &= a(r_0) \sin (\theta/\lambda v_0), \end{aligned} \tag{22}$$

where $a(r_0)$ [>0] denotes the ultimate velocity magnitude assignable arbitrarily upon and constant along each individual streamline. The *density distribution* ρ, can be computed by recourse to the basic relation (valid for ideal gases in absence of body forces)

$$\rho = \frac{\gamma p M^2}{V^2} = \frac{\gamma p}{a^2(r_0) c^2},$$

where p and c^2 are given by (15) and (18).

Examples of flow patterns. The general nature of the flow patterns associated with several choices of v_0 is indicated in Figs. 1 to 4. These patterns were computed from equation (12) using $\lambda = 6$ (the value for air) and values for v_0 chosen to give limiting values of θ of $\pi/4$, $\pi/2$, π, and $3\pi/2$ (that is, flows turning from $\theta = 0$ through $-\pi/4$, 0, $\pi/2$, and π). The flow of Fig. 3 is of particular interest, representing a flow around the edge of an infinitely thin flat plate. This flow is similar to the irrotational Ringleb flow, but lacks the limiting lines which are characteristic of the latter.

UPPER AND LOWER BOUNDS FOR THE SOLUTION OF THE FIRST BOUNDARY VALUE PROBLEM OF ELASTICITY*

By J. B. DIAZ and H. J. GREENBERG (*Brown University*)

In another paper,[1] we have considered the boundary value problem

$$\Delta\Delta w = p, \qquad \text{in } R,$$
$$w = f, \qquad \frac{\partial w}{\partial n} = g, \qquad \text{on } C, \tag{1}$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$, R is a plane domain with boundary C, and $\partial/\partial n$ denotes differentiation in the direction of the outer normal to C. Assuming the existence and uniqueness of the solution of (1), we developed a method for obtaining upper and lower bounds for $w(x_0, y_0)$, the point x_0, y_0 of R being given in advance. The upper and lower bounds were given in terms of integrals of two classes of functions, one satisfying certain boundary conditions and the other certain partial differential equations.

In the present paper we shall be concerned with the application of the same method to a boundary value problem involving a system of partial differential equations. For definiteness we shall deal only with the first boundary value problem of three dimensional elasticity, i.e.

$$\mu w_{i,jj} + (\lambda + \mu) w_{j,ji} + X_i = 0, \qquad \text{in } R,$$
$$w_i = \text{a given function}, \qquad \text{on } S, \tag{2}$$

$(i, j = 1, 2, 3)$, where R is a domain in three dimensions bounded by a surface S, w_i is the i-th component of the displacement, λ and μ are Lamé's constants, X_i is the i-th component of the body force, commas indicate partial differentiation, and a repeated subscript indicates summation over the range 1, 2, 3. The differential equations in (2) are known as Navier's equations.

Given a point (ξ_1, ξ_2, ξ_3) of R, we seek to obtain upper and lower bounds for the numbers $w_1(\xi_1, \xi_2, \xi_3)$, $w_2(\xi_1, \xi_2, \xi_3)$, and $w_3(\xi_1, \xi_2, \xi_3)$. The "Green's formulas" needed in the discussion are included in the first section. The second section deals with some auxiliary inequalities and the third with the upper and lower bounds sought. A comparison with the paper quoted in footnote 1 reveals that the difficulty involved in the transition from one equation to a system of equations is principally one of notation.

1. Green's identities.[2] In deriving Green's identities for the system (2), we shall save space by using a convenient notation. We shall write $\phi = (\phi_1, \phi_2, \phi_3)$, for example, where the ϕ_i are real functions defined on R. Also, we shall need Green's theorem

$$\int_R A_{,i}\, dR = \int_S A n_i\, dS, \tag{3}$$

---

*Received Feb. 18, 1948. This paper is based on a report prepared for Watertown Arsenal under a contract in Applied Mechanics.

[1] J. B. Diaz and H. J. Greenberg, *Upper and lower bounds for the first biharmonic boundary value problem*, to appear in J. Math. Phys.

[2] See O. D. Kellogg, *Potential theory*, 1929. The formulas (6), (7), and (9) of this paper correspond to those for Laplace's equation which appear on pages 212, 215, and 219, respectively, of Kellogg's book.

$j = 1, 2, 3$, where n_j is the j-th component of the outer normal to S.

Let $\phi = (\phi_1 , \phi_2 , \phi_3)$ and $\psi = (\psi_1 , \psi_2 , \psi_3)$. From Green's theorem we have at once that

$$\int_R [\mu(\phi_i\psi_{i,j})_{,j} + (\lambda + \mu)(\phi_i\psi_{j,j})_{,i}]\, dR = \int_S \phi_i[\mu\psi_{i,j}n_j + (\lambda + \mu)\psi_{j,j}n_i]\, dS. \quad (4)$$

Introducing the notations

$$(\phi, \psi) = \int_R [\mu\phi_{i,j}\psi_{i,j} + (\lambda + \mu)\phi_{i,i}\psi_{j,j}]\, dR,$$

$$(\phi, \psi)_R = \int_R \phi_i[\mu\psi_{i,jj} + (\lambda + \mu)\psi_{j,ji}]\, dR, \quad (5)$$

$$(\phi, \psi)_S = \int_S \phi_i[\mu\psi_{i,j}n_j + (\lambda + \mu)\psi_{j,j}n_i]\, dS,$$

we may rewrite Eq. (4) as follows:

$$(\phi, \psi)_R + (\phi, \psi) = (\phi, \psi)_S \quad (6)$$

The integral $(\phi, \psi) = (\psi, \phi)$ plays the role of the Dirichlet integral $\int_R (u_x v_x + u_y v_y + u_z v_z)\, dR$ in the theory of Laplace's equation. Interchanging ϕ and ψ in (6), and subtracting the resulting equation from (6), we obtain Green's reciprocal theorem

$$(\phi, \psi)_R - (\psi, \phi)_R = (\phi, \psi)_S - (\psi, \phi)_S . \quad (7)$$

In order to obtain the remaining formula of Green, we must employ the "singular" solutions of Navier's equations (2) with body forces zero, which were given by Lord Kelvin. These three solutions give the displacements produced throughout space by a concentrated unit force in the x_1 , x_2 , and x_3 directions respectively. The solutions $Z^{(k)}$, $k = 1, 2, 3$, of the homogeneous system (2), with (ξ_1 , ξ_2 , ξ_3) as singular point, are given by

$$Z_i^{(k)} = -\frac{\lambda + \mu}{8\pi\mu(\lambda + 2\mu)}\left[r_{,ki} - \frac{\lambda + 2\mu}{\lambda + \mu} r_{,ll}\delta_{ki}\right], \quad (8)$$

$i = 1, 2, 3$, where r is the distance from (ξ_1 , ξ_2 , ξ_3) to (x_1 , x_2 , x_3), and δ_{ki} is Kronecker's delta (unity if $i = k$ and zero if $i \neq k$). If we now employ a procedure common in potential theory, replacing ψ by $Z^{(k)}$ in (7), the integration being extended over the common part of R and the exterior of a sufficiently small sphere with center at (ξ_1 , ξ_2 , ξ_3), and pass to the limit as the sphere shrinks toward (ξ_1 , ξ_2 , ξ_3), we obtain[3]

$$\phi_k(\xi_1 , \xi_2 , \xi_3) = -(Z^{(k)}, \phi)_R + (Z^{(k)}, \phi)_S - (\phi, Z^{(k)})_S . \quad (9)$$

2. Preliminary inequalities. Starting with (ϕ, ψ), defined in (5), and writing $\alpha\phi = (\alpha\phi_1 , \alpha\phi_2 , \alpha\phi_3)$, for real α, we have

$$(\alpha\phi + \beta\psi, \alpha\phi + \beta\psi) = \alpha^2(\phi, \phi) + 2\alpha\beta(\phi, \psi) + \beta^2(\psi, \psi) \geqq 0,$$

[3]See A. E. H. Love, *Elasticity*, New York, 1944, p. 245, and C. Somigliana, Annali di Matematica (2) 17, 41 (1889).

for every real α and β, which yields Schwarz' inequality

$$[(\phi, \psi)]^2 \leqq (\phi, \phi) \cdot (\psi, \psi). \tag{10}$$

Let $k = 1, 2, 3$ be fixed, and consider the solution $W^{(k)} = (W_1^{(k)}, W_2^{(k)}, W_3^{(k)})$ of the boundary value problem

$$\begin{aligned} \mu W_{i,jj}^{(k)} + (\lambda + \mu) W_{j,ji}^{(k)} &= 0, \qquad \text{in } R, \\ W_i^{(k)} &= -Z_i^{(k)}, \qquad \text{on } S, \end{aligned} \tag{11}$$

$i = 1, 2, 3$. ($W^{(k)}$ is the "regular part" of a certain Green's function). Together with $W^{(k)}$ we shall consider "vectors" $V = (V_1, V_2, V_3)$ and $U^{(k)} = (U_1^{(k)}, U_2^{(k)}, U_3^{(k)})$ such that

$$\mu V_{i,jj} + (\lambda + \mu) V_{j,ji} = 0, \qquad \text{in } R, \tag{12}$$

$i = 1, 2, 3$, and

$$U_i^{(k)} = W_i^{(k)} = -Z_i^{(k)}, \qquad \text{on } S, \tag{13}$$

$i = 1, 2, 3$, respectively. In connection with the vector w of (2) we shall introduce the vectors v and u such that[4]

$$\mu v_{i,jj} + (\lambda + \mu) v_{j,ji} + X_i = 0, \qquad \text{in } R, \tag{14}$$

$i = 1, 2, 3$, and

$$u_i = w_i, \qquad \text{on } S. \tag{15}$$

It is now easy to verify that

$$\begin{aligned} (u - w, u - w) \\ (v - w, v - w) \end{aligned} \leqq (u - v, u - v). \tag{16}$$

Recalling the definitions of the vectors involved, and using (5) and (6), we have

$$\begin{aligned} (u - v, u - v) &= ((u - w) + (w - v), (u - w) + (w - v)) \\ &= (u - w, u - w) + (v - w, v - w), \end{aligned} \tag{17}$$

since $u - w$ vanishes on S and $v - w$ satisfies the homogeneous system of equations in R.

The two known minimum principles for the solution of (2) can be immediately deduced from the inequalities (16), as will be seen in Sec. 4.

3. Upper and lower bounds for w at a given point. Applying (9) to the vector w we obtain

$$w_k(\xi_1, \xi_2, \xi_3) = -(Z^{(k)}, w)_R + (Z^{(k)}, w)_S - (w, Z^{(k)})_S, \tag{18}$$

$k = 1, 2, 3$, and we are seeking bounds on the left hand members of (18). Of the three integrals in (18), only $(Z^{(k)}, w)_S$ cannot be computed directly from the conditions (2)

[4] C. Somigliana, Annali di Matematica, Ser. 2, vol. 17, 1889, p. 39, gives a method for obtaining solutions of (12) and (14). Let the functions $F_l (l = 1, 2, 3,)$ satisfy $\Delta\Delta F_l = -X_l/\mu$. Then the functions $v_i = -(\lambda + \mu)/(\lambda + 2\mu) F_{l,li} + F_{i,ll}$ $(i = 1, 2, 3)$, satisfy (14). In particular, the singular solution $Z^{(k)}$, given in (8), is obtained by taking $F_l = r\delta_{kl}$.

on w, and hence it is only this integral for which we have to find bounds. For each k, Schwarz' inequality, (10), yields

$$[(u - w,\ U^{(k)} - W^{(k)})]^2 \leqq (u - w,\ u - w)\cdot(U^{(k)} - W^{(k)},\ U^{(k)} - W^{(k)}). \tag{19}$$

This last equation, together with (16), implies that

$$[(u - w,\ U^{(k)} - W^{(k)})]^2 \leqq (u - v,\ u - v)\cdot(U^{(k)} - V,\ U^{(k)} - V). \tag{20}$$

But, from (6) we have that

$$\begin{aligned}(u - w,\ U^{(k)} - W^{(k)}) &= (u,\ U^{(k)}) - (U^{(k)},\ w) - (u - w,\ W^{(k)}) \\ &= (u,\ U^{(k)}) + (U^{(k)},\ w)_R - (U^{(k)},\ w)_S\ .\end{aligned} \tag{21}$$

From (13),

$$-(U^{(k)},\ w)_S = (Z^{(k)},\ w)_S\ ,$$

the "unknown" surface integral of (18). Substituting for this surface integral from (18), and combining the resulting equation with (20), we obtain

$$(w_k(\xi_1\ ,\ \xi_2\ ,\ \xi_3) - b_k)^2 \leqq a\cdot c_k\ , \tag{22}$$

where

$$\left.\begin{aligned} a &= (u - v,\ u - v), \\ c_k &= (U^{(k)} - V,\ U^{(k)} - V), \\ b_k &= -(u,\ U^{(k)}) - (U^{(k)},\ w)_R \\ &\quad -(Z^{(k)},\ w)_R - (w,\ Z^{(k)})_S\ . \end{aligned}\right\} \tag{23}$$

Given $k = 1$, 2, or 3, and four vectors u, v, $U^{(k)}$, and V, satisfying conditions (15), (14), (13), and (12), respectively, equations (22) and (23) yield upper and lower bounds for $w_k(\xi_1\ ,\ \xi_2\ ,\ \xi_3)$.

Another set of bounds can be obtained by starting with

$$[(v - w,\ V - W^{(k)})]^2 \leqq (v - w,\ v - w)\cdot(V - W^{(k)},\ V - W^{(k)}), \tag{24}$$

instead of (19). From (16) we have that

$$[(v - w,\ V - W^{(k)})]^2 \leqq (u - v,\ u - v)\cdot(U^{(k)} - V,\ U^{(k)} - V), \tag{25}$$

and from (6) it follows that

$$\begin{aligned}(v - w,\ V - W^{(k)}) &= (v,\ V) + (W^{(k)},\ w - v) - (w,\ V) \\ &= (v,\ V) + (W^{(k)},\ w - v)_S - (w,\ V)_S\ .\end{aligned} \tag{26}$$

The "unknown" surface integral of (18) appears, with a minus sign, on the right hand side of (26). Substituting for this surface integral from (18), and combining the resulting equation with (25), we obtain

$$(w_k(\xi_1\ ,\ \xi_2\ ,\ \xi_3) - b_k')^2 \leqq a\cdot c_k\ , \tag{27}$$

where

$$\left.\begin{aligned} a &= (u - v,\ u - v), \\ c_k &= (U^{(k)} - V,\ U^{(k)} - V), \\ b_k' &= (v,\ V) - (W^{(k)},\ v)_S - (w,\ V)_S \\ &\quad -(Z^{(k)},\ w)_R - (w,\ Z^{(k)})_S\ . \end{aligned}\right\} \tag{28}$$

Again, given $k = 1, 2$, or 3, and four vectors $u, v, U^{(k)}$, and V, satisfying conditions (15), (14), (13), and (12), respectively, equations (27) and (28) yield upper and lower bounds for $w_k(\xi_1\ ,\ \xi_2\ ,\ \xi_3)$. The relations (22), (23), (27), and (28) are the inequalities we wished to derive.

It is clear[5] that an iteration procedure for improving the upper and lower bounds already obtained may be given in terms of sequences u_i and w_i of vectors satisfying the homogeneous boundary conditions

$$u_{i1} = u_{i2} = u_{i3} = 0, \qquad \text{on } S,$$

and the homogeneous system of equations (12) respectively.

4. Minimum principles for w. It is interesting to observe that it is possible to deduce from (16) two known minimum principles associated with the solution w of (2). One principle picks out w from among the u's (the vectors satisfying the boundary conditions of (2)), and the other principle singles out w from among the v's (the vectors satisfying the system of partial differential equations of (2)).

Let us derive the first minimum principle ("principle of minimum potential energy"). From (5), $(\phi, \phi) = 0$ if and only if ϕ is a constant vector, i.e., each of $\phi_1\ ,\ \phi_2\ ,\ \phi_3$ reduces to a constant. In view of this remark, equations (16) and (17) imply that, for any fixed v,

$$(v - w,\ v - w) \leqq (u - v,\ u - v), \tag{29}$$

the equality sign holding if and only if $u = w$. Expanding (29), employing (6) and (15), yields

$$(w,\ w) + 2(w,\ v)_R \leqq (u,\ u) + 2(u,\ v)_R\ . \tag{30}$$

Taking (5) and (14) into account, this result may be stated as follows: the functional

$$\int_R [\mu u_{i,j}u_{i,j} + (\lambda + \mu)u_{i,i}u_{j,j} - 2u_iX_i]\ dR \tag{31}$$

is minimized, over the class of vectors u satisfying (15), by the solution w of (2).

The second minimum principle ("principle of minimum complementary energy", or "Castigliano's principle") may be derived in a similar manner. Equations (16) and (17) imply that, for any fixed u,

$$(u - w,\ u - w) \leqq (u - v,\ u - v), \tag{32}$$

the equality sign holding if and only if $v - w$ is a constant vector. With the aid of (6) and (14), Eq. (32) yields

[5] See, for instance, section 3 of the paper quoted in footnote 1.

$$(w, w) - 2(u, w)_S \leqq (v, v) - 2(u, v)_S . \tag{33}$$

Taking (5) and (15) into account, this result may be stated as follows: the functional

$$\int_R [\mu v_{i,j} v_{i,j} + (\lambda + \mu) v_{i,i} v_{j,j}]\, dR - 2 \int_S w_i [\mu v_{i,j} n_j + (\lambda + \mu) v_{j,j} n_i]\, dS, \tag{34}$$

is minimized, over the class of all vectors v satisfying (14), by $w + c$, where w is the solution of (2) and c is any constant vector.

ON SOME SINGULAR SOLUTIONS OF THE TRICOMI EQUATION*

By G. F. CARRIER and F. E. EHLERS (*Brown University*)

1. Introduction. In this note we intend to develop briefly some singular solutions of the Tricomi equation. These solutions have application in the hodograph techniques for the theory of compressible fluids. The flows to which they apply will be discussed in a later work but it is felt that the singular solutions are of sufficient interest to merit this presentation.

2. The singular solutions. We shall consider the equations

$$\varphi_x = \psi_y , \tag{1}$$

$$\psi_x = -\varphi_y / y \tag{2}$$

which imply

$$\psi_{yy} + y\psi_{xx} = 0 \tag{3}$$

and

$$\varphi_{xx} + (\varphi_y / y)_y = 0. \tag{4}$$

Equation (3) is the Tricomi equation. The solution of primary interest has the following properties:

$$\psi_x \sim 1/y \qquad \text{on } x = 0,\ y \to 0;$$

$$\varphi_y \equiv 0 \qquad \text{on } y = 0, \text{ for } x \neq 0;$$

$$\psi, \varphi, \text{ regular in } y > 0 \text{ and on } x = 0,\ y \neq 0.$$

It is evident that if φ, ψ, are not regular in $y < 0$, the branch lines will occur along the characteristics $x^2 + 4y^3/9 = 0$. The development that follows is strictly formal and the proof that the solutions are those sought is readily found by substituting them into the original equations. To find them, we first replace y and φ by

$$s = (2/3)y^{3/2}, \qquad \varphi = (2/3)^{2/3} y \chi(x, s) = s^{2/3} \chi(x, s).$$

*Received May 14, 1948.

Equation (4) becomes

$$(s\chi_s)_s - 4\chi/9s + s\chi_{xx} = 0. \tag{5}$$

We may obtain a solution valid in the elliptic domain ($y > 0$) by the classical Fourier transform technique. We define

$$\bar{\chi}(\xi, \alpha) = \int_{-\infty}^{\infty}\int_{0}^{\infty} sJ_{2/3}(\alpha s)e^{-i\xi x}\chi(x, s)\, ds\, dx \tag{6}$$

and multiply Eq. (5) through by $J_{2/3}(\alpha s)e^{-i\xi x}$. Integration over the domain $y > 0$ of this equation can be carried out as follows. Integration by parts with regard to s yields contributions including some functions of x arising from the boundary condition on $y = 0$. The only admissible contribution of this sort is a delta function at the origin, i.e., $\delta(x - 0)$. Subsequent integration over x transforms Eq. (5) into the form

$$(\xi^2 + \alpha^2)\bar{\chi} = 1. \tag{7}$$

However, as implied by Eq. (6),

$$\chi = \frac{1}{4\pi^2}\int_{0}^{\infty}\int_{-\infty}^{\infty} e^{i\xi x}J_{2/3}(\alpha s)\bar{\chi}(\xi, \alpha)\, d\alpha\, d\xi \tag{8}$$

which reduces readily to

$$\chi(x, s) = C\int_{0}^{\infty} e^{-\alpha|x|}J_{2/3}(\alpha s)\, d\alpha, \qquad C \text{ being a constant.} \tag{9}$$

It is also evident that $\int^x \chi\, dx$ or χ_x is a solution of Eq. (5) and, in fact, it can be verified by a simple substitution that

$$\chi^{(\mu)}(x, s) = \int_{0}^{\infty} \alpha^{\mu-1}e^{-\alpha|x|}J_{2/3}(\alpha s)\, d\alpha \tag{10}$$

is a solution. However[1], the integral of Eq. (10) is (when it exists)

$$\chi^{(\mu)}(x, s) = \frac{(s/2)^{2/3}\Gamma(\mu + 2/3)}{\Gamma(5/3)(x^2 + s^2)^{\mu/2+1/3}}\, F\left(\frac{3\mu + 2}{6}, \frac{4 - 3\mu}{6}; \frac{5}{3}; \frac{s^2}{x^2 + s^2}\right). \tag{11}$$

It is evident that (in Eq. 8) we might have used $J_{-2/3}(\alpha s)$ instead of the Bessel function of order 2/3. In this case

$$\chi^{(\mu')}(x, s) = \frac{(s/2)^{-2/3}\Gamma(\mu - 2/3)}{\Gamma(1/3)(x^2 + s^2)^{\mu/2-1/3}}\, F\left(\frac{3\mu - 2}{6}, \frac{1 - 3\mu}{6}; \frac{1}{3}; \frac{s^2}{x^2 + s^2}\right). \tag{12}$$

It is now readily shown that a linear combination of the functions defined by Eqs. (11), (12), is the desired solution provided $\mu = 0$. This again is a formal step since the integral which leads to Eq. (12) does not converge for $\mu = 0$. However, by arguments involving analytic continuation (or again by the trivial substitution) the functions of (11) and (12) can be seen to constitute proper solutions for all μ.

For $\varphi^{(\mu)}$ we choose the combination

$$\varphi^{(\mu)} = A(\mu)s^{2/3}\chi^{(\mu)} + B(\mu)s^{2/3}\chi^{(\mu')}, \tag{13}$$

[1]See G. N. Watson, *Bessel functions*, Macmillan Company, New York, 1944, p. 385.

where

$$A(\mu) = \frac{\Gamma(1/3)\Gamma(2/3)\Gamma(1/2)}{\Gamma(\mu - 2/3)\Gamma(\mu + 1/3)\Gamma(5/6 - \mu/2)},$$

$$B(\mu) = \frac{\Gamma(-2/3)\Gamma(5/3)\Gamma(1/2)}{\Gamma(1/6 - \mu/2)\Gamma(\mu/2 - 1/3)\Gamma(\mu + 2/3)}.$$

Any other combination is not regular in a sufficiently general region. With the aid of an identity[2] for the hypergeometric functions and of Eq. (11) and (12), this reduces to the following equation for $\mu = 0$

$$\varphi = (x^2 + s^2)^{1/3} F\left(-\frac{1}{3}, \frac{1}{6}; \frac{1}{2}; \frac{x^2}{x^2 + s^2}\right). \tag{14}$$

This is the desired solution for φ.

Retaining the integrals in Eq. (13), we compute the conjugate function ψ for arbitrary y. Since

$$\psi_x = -(1/y)(\partial\varphi/\partial y) = -(y^{-1/2})(\partial\varphi/\partial s) = -(2/3)^{1/3} s^{-1/3} \partial\varphi/\partial s$$

we obtain

$$(3/2)^{1/3}\psi = 2^{-2/3} A(\mu) s^{1/3} \int_0^\infty e^{-|x|\alpha} J_{1/3}(\alpha s)\alpha^{\mu-1}\, d\alpha$$

$$- 2^{2/3} B(\mu) s^{1/3} \int_0^\infty e^{-|x|\alpha} J_{-1/3}(\alpha s)\alpha^{\mu-1}\, d\alpha.$$

Replacing the integrals and simplifying the coefficients leads to

$$\psi = \frac{4^{1/3}}{3} \frac{y}{(x^2 + 4y^3/9)^{1/6}} F\left(\frac{1}{6}, \frac{2}{3}; \frac{4}{3}; \frac{4y^3/9}{x^2 + 4y^3/9}\right)$$

$$- (4/3)^{1/3} (x^2 + 4y^2/9)^{1/6} F\left(-\frac{1}{6}, \frac{1}{3}; \frac{2}{3}; \frac{4y^3/9}{x^2 + 4y^3/9}\right) \tag{15}$$

for $y = 0$. It is easily verified that (14) and (15) define functions having the desired properties.

One should note that these functions can be taken as single valued except in the domain $(x^2 + s^2) < 0$. In this region they are essentially triple valued and, in fact, this is precisely the property required for the application hinted at in the introduction.[3] We might add that a more general solution regular in $x^2 + s^2 > 0$, multi-valued in $x^2 + s^2 < 0$, and having the behavior $\varphi_s \sim y^{3\nu-1}$ as $y \to 0$, $x \equiv 0$, is

$$\varphi = (x^2 + s^2)^\nu F\left(-\nu, \nu - \frac{1}{6}; \frac{1}{2}; x^2/(x^2 + s^2)\right). \tag{16}$$

It seems worth while to also record another singular solution of interest. It is the

[2] E. T. Whitaker and G. N. Watson, *Modern analysis*, University Press, Cambridge, 1940, p. 291.

[3] A discussion of the necessity of this property can be found in Lighthill, M. J., *The hodograph transformation in trans-sonic flow. I. Symmetrical channels*, Proc. Roy. Soc. London (A) 191, 323 (1947).

solution such that φ has a logarithmic singularity at the point x_0 , y_0 ; ψ (of course) has an "arc tan behavior" at this point. It is immediately evident that we are looking for the Green's function of Eq. (5), say for the region $y > 0$. To find this, we merely put a spurious non-homogeneous term $N(x, s)$ on the right side of (5), define $\bar{N}$ as we did $\bar{\chi}$, and operate as before.

This time we find

$$(\xi^2 + \alpha^2)\bar{\chi} = \bar{N} \tag{17}$$

if we allow the boundary contributions to vanish. The inversion of Eq. (17) leads to

$$\chi = \int_{-\infty}^{\infty}\int_{0}^{\infty} s_0 N(x_0 , s_0) \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{\alpha e^{i\xi(x-x_0)} J_{2/3}(\alpha s)}{\xi^2 + \alpha^2} J_{2/3}(\alpha s_0)\, d\xi\, d\alpha\, dx_0\, ds_0 \tag{18}$$

where the double integral over ξ and α must define the Green's function. Thus, we write

$$\begin{aligned} G(x, x_0 ; s, s_0) &= \int_0^{\infty} e^{-\alpha|x-x_0|} J_{2/3}(\alpha s) J_{2/3}(\alpha s_0)\, d\alpha \\ &= A s^{-1/2} Q_{1/6}\left(\frac{(x - x_0)^2 + s^2 + s_0^2}{2ss_0}\right) \end{aligned} \tag{19}$$

where A is some constant.

The corresponding solutions φ and ψ are

$$\varphi = y^{1/4} Q_{1/6}\left(\frac{(x - x_0)^2 + y^3 + y_0^3}{2(yy_0)^{3/2}}\right) \tag{20}$$

$$\psi = \int^x (\varphi_v / y)\, dx \tag{21}$$

It is interesting to note that the stream function ψ is closely related to one found by Weinstein.[4]

If we had formally used $J_{-2/3}$ again, the index $1/6$ appearing in the Legendre function would be replaced by $-7/6$. Either solution may readily be continued into the hyperbolic domain.

THE LOST SOLUTIONS IN AXIALLY SYMMETRIC IRROTATIONAL COMPRESSIBLE FLUID FLOW*

By H. J. STEWART (*California Institute of Technology*)

1. Since the two dimensional potential equation for the irrotational flow of a compressible fluid may be linearized by a Legendre transformation from the physical plane to the velocity (hodograph) plane, this linearizing transformation is usually one of the first steps taken in the study of two dimensional compressible fluid flows. This trans-

[4]Weinstein, A., *On axially symmetric flows*, Q. Appl. Math. **5**, 429 (1948).

*Received Jan. 12, 1948.

formation is degenerate for any flow for which one velocity component is a function of the other velocity component. In this case the entire physical plane maps into a line in the hodograph plane. Such a solution to the potential flow equation is called a "lost" solution because the Legendre transformation to the hodograph plane is not valid in these circumstances and cannot be used to investigate such a flow. In the two dimensional supersonic theory this type of flow, generally called Prandtl-Meyer flow,[1] is of great importance since it includes the expanding flow around a corner or over any convexly curved surface. It is also important in the design of supersonic wind tunnel nozzles since the matching flow between a two dimensional source flow (representing the initial expansion near the nozzle throat) and the rectilinear flow in the working section is of the Prandtl-Meyer type.

Because of the importance of the lost solution in the two dimensional case, the question of the existence and nature of similarly degenerate flows in three dimensional flow problems is of considerable interest. The complete solution to this question for one class of three dimensional flows, axially symmetric flows, can be obtained by a very simple extension of a method used by Bateman[2] in treating the two dimensional problem.

2. Consider an axially symmetric irrotational compressible fluid flow with x being the distance along the axis of symmetry and y being the radial distance from the axis. Let u and v be the velocity components in the x and y directions respectively, and let ρ be the density of the fluid. The functional relation between the velocity components may be written in a parametric form as

$$u = A(s), \qquad v = B(s), \qquad \rho = F(s), \tag{1}$$

where $F(s)$ may be expressed in terms of $A(s)$ and $B(s)$ by the momentum integral and the parameter $s = s(x, y)$ represents the lines of constant velocity in the physical plane. Since the flow is irrotational,

$$\frac{\partial v}{\partial x} = \frac{\partial u}{\partial y} \tag{2}$$

or

$$B'(s)\frac{\partial s}{\partial x} = A'(s)\frac{\partial s}{\partial y}. \tag{3}$$

This shows that the lines of constant s must be straight lines of the type

$$xA'(s) + yB'(s) = C'(s), \tag{4}$$

where $C'(s)$ is an arbitrary continuous function. The velocity potential φ is thus

$$\varphi = xA(s) + yB(s) - C(s). \tag{5}$$

The functions $A(s)$ and $B(s)$ must be so related that the continuity equation is satisfied. The continuity equation for axially symmetric flow is

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) + \frac{\rho v}{y} = 0. \tag{6}$$

---

[1] W. F. Durand, *Aerodynamic theory*, vol. 3. J. Springer, Berlin, 1935, p. 243.

[2] H. Bateman, *Irrotational motion of a compressible inviscid fluid*, Proc. Nat. Acad. Sci., **16**, 816 (1930).

Thus

$$(FA' + AF')\frac{\partial s}{\partial x} + (FB' + BF')\frac{\partial s}{\partial y} + \frac{FB}{y} = 0. \tag{7}$$

By Eq. (3), this may be written

$$\frac{FA' + AF'}{B'} + \frac{FB' + BF'}{A'} = -\frac{FB}{B'y\ \partial s/\partial x} = -\frac{FB}{A'y\ \partial s/\partial y} \tag{8}$$

unless $B'(s)\ \partial s/\partial x = A'(s)\ \partial s/\partial y = 0$. This special case will be discussed later. Except in this special case, Eq. (8) shows that a solution of the type indicated by Eq. (1) is possible only if

$$y\frac{\partial s}{\partial x} = g(s), \qquad y\frac{\partial s}{\partial y} = h(s), \tag{9}$$

where $g(s)$ and $h(s)$ are arbitrary but related continuous functions of s.

The functions $s(x, y)$ which satisfy Eq. (9) can be determined by inverting variables and considering $x = x(s, y)$. Then Eq. (9) becomes (if $\partial s/\partial x \neq 0$)

$$\frac{\partial x}{\partial s} = \frac{1}{\partial s/\partial x} = \frac{y}{g(s)} = yG(s), \tag{10}$$

$$\frac{\partial x}{\partial y} = -\frac{\partial s/\partial y}{\partial s/\partial x} = -\frac{h(s)}{g(s)} = H(s).$$

These expressions may be simply integrated, and it is seen that

$$x = yH(s) + c, \tag{11}$$

where c is a constant and $G(s) = H'(s)$. Thus the only function $s(x, y)$ which satisfies Eq. (9) is

$$s = s\left(\frac{y}{x - c}\right). \tag{12}$$

The excepted case, $\partial s/\partial x = 0$, is given by the limit as $c \to \infty$. If c is finite it may always be made zero by a proper choice of the origin. There are thus only two solutions of Eq. (9),

$$s = s(y), \tag{13}$$

$$s = s(y/x), \tag{14}$$

The derivation of Eq. (9) breaks down if both sides of Eq. (3) are identically zero. The simplest case occurs if $A'(s) = B'(s) = 0$. This corresponds to a uniform rectilinear flow which is axially symmetric if $B = 0$. This is obviously the simplest lost solution and is contained in both Eq. (13) and (14) as $s = s_0$, a constant. If $s = s(x)$, both sides of Eq. (3) are identically zero if $B'(s) = 0$. The continuity equation can then be satisfied only if $B = 0$; so this, too, corresponds to a uniform flow. The case, $\partial s/\partial x = 0$ and $A'(s) \neq 0$, is the limiting case of Eq. (13). This case includes, beside the rectilinear flow along the axis, a two dimension line source on the axis of symmetry or a superposition of the two flows. The remaining case, Eq. (14), leads by Eq. (5) to the Taylor-Maccoll[3]

[3]G. I. Taylor and J. W. Maccoll, *Air pressure on a cone moving at high speeds*, Proc. Roy. Soc. (A) 139, 278 (1933).

potential equation and corresponds to the supersonic flow past a cone in the range of Mach numbers for which the shock wave at the nose is attached to the cone. The only lost solutions of the axially symmetric potential flow equation are thus seen to be a rectilinear flow along the axis, a two dimensional line source on the axis, and the Taylor Maccoll flow past a cone. The first two flows may be superimposed, but the result is of little interest. Only the third conical flow is an essentially three dimensional flow pattern.

In conclusion it can be stated that all the lost solutions of axially symmetric irrotational compressible flow are already well known solutions. This result was apparently known to Bateman; however, it does not appear in any of his writings. A discussion of another class of three dimensional lost solutions, conical flows, has been given by Busemann.[4]

A GENERAL STABILITY CRITERION FOR LINEAR OSCILLATING SYSTEMS WITH CONSTANT TIME LAG*

By H. I. ANSOFF and J. A. KRUMHANSL (*Brown University*)

Several formulations[1] have recently been given for the stability criterion of a lumped-parameter linear oscillating system with constant time lag given by the equation

$$Iy''(t) + Ry'(t) + Ky(t) = -Sy'(t - \tau), \tag{1}$$

where I, R, K, S and τ are real positive constants and $Sy'(t - \tau)$ is the feedback term.

By applying Cauchy's index theorem to the operational form of (1) it is possible to generalize the discussion to feedback proportional to any derivative of the dependent variable taken at time $t - \tau$; this will be called a retarded derivative. It is found that the resulting stability criterion can be written in an easily computed form and that the permissible range in the time lag τ can be stated explicitly.

Equation (1) is generalized to

$$Iy''(t) + Ry'(t) + Ky(t) = -Sy^{(n)}(t - \tau), \tag{2}$$

where n is allowed any integer positive value. Letting $F(p)$ be the Laplace transform of $y(t)$, one obtains[2]

$$F(p) = \frac{L(p)}{Y(p) + Sp^n e^{-p\tau}} = \frac{L(p)}{Y(p)} \frac{1}{1 + Sp^n e^{-p\tau}/Y(p)}, \tag{3}$$

where

$$\begin{aligned} L(p) &= Ipy(0) + Iy'(0) + Ry(0), \\ Y(p) &= Ip^2 + Rp + K. \end{aligned} \tag{4}$$

---

[4] A. Busemann, *Aerodynamischer Auftrieb bei Überschallgeschwindigkeit*, Luftfahrtforschung, **12**, 210 (1935).

*Received Feb. 27, 1948.

[1] See references [1] and [2]. In the following, numbers in square brackets refer to the Bibliography at the end of this paper.

[2] The Laplace transform is defined here as in [3]. For other examples of this procedure see reference [4].

By Mellin's inversion theorem

$$y(t) = \frac{1}{2\pi i}\int_{\Gamma} \frac{L(p)}{Y(p)} \frac{e^{pt}\,dp}{1 + Sp^n e^{-p\tau}/Y(p)} \qquad \text{for} \qquad t > 0, \tag{5}$$

where Γ is a contour enclosing all of the poles of the integrand in the p-plane.

Evaluation of this integral by the method of residues yields an expression of the form

$$y(t) = \sum_0^n A_n \exp(p_n t), \tag{6}$$

where p_n are the poles of the integrand of (5). It is easily shown [5] that in general there is an infinity of these poles. Among these poles, however, only p_n with positive real parts will cause (6) to be unstable. The subsequent investigation is therefore devoted to a formulation of conditions under which the system has no p_n with positive real parts.

The functions $L(p)$ and e^{pt} in (5) have no poles in the finite p plane, and $Y(p)$ has no roots in the right half of the p plane, as shown by (4). Thus a necessary and sufficient condition for the stability of (6) is that

$$1 + S\frac{p^n e^{-p\tau}}{Y(p)} = 0 \tag{7}$$

have no roots in the right half of the p plane.

In order to determine the roots of (7), we introduce the Cauchy index theorem [6] which may be simplified as follows:

Hypothesis. 1) $w = f(p)$ is an analytic function in a simply connected domain D bounded by a contour β, and

2) $f(p) \neq 0$ for p on β.

Conclusion. If p traverses β in a counter-clockwise direction, then w will traverse a closed curve in the w-plane. Further, the *number of zeros of $f(p)$ in D is equal to the net number of times the contour in the w-plane encircles the origin.*

For β we choose a Bromwich contour in the p plane consisting of the imaginary axis[3] and an infinite semi-circle in the right half of the plane with center at the origin.

Applying the theorem to (7), we note that the number of roots of (7) with positive real parts is equal to the number of times the graph of the function

$$f(p) = \frac{Sp^n e^{-p\tau}}{Ip^2 + Rp + K} \tag{8}$$

encloses the point $(-1, 0)$ as p traverses β. It can be shown [5] that $f(p) \to 0$ for all values of n, as the radius R of the semi-circular part of β tends to ∞. It remains to examine $f(p)$ on the imaginary axis.

If we let $p = i\alpha$, where α is real, Eq. (8) can be written

$$f(i\alpha) = \frac{S\alpha^n e^{i\beta}}{[(K - I\alpha^2)^2 + (R\alpha)^2]^{\frac{1}{2}}}, \tag{9a}$$

where

[3] It is shown in [5] that roots *on* the imaginary axis occur only for a very critical combination of the parameters of the system.

$$\beta = -\alpha\tau + \phi + n\pi/2, \tag{9b}$$

$$\tan \phi = \frac{R\alpha}{K - I\alpha^2}, \tag{9c}$$

and n corresponds to the order of the feedback term

The graph of (9) for any fixed n is a sensed spiral with a variable amplitude given by $|f(i\alpha)|$. Figure 1 shows typical graphs for $n = 0, 1, 2$, respectively, in the range

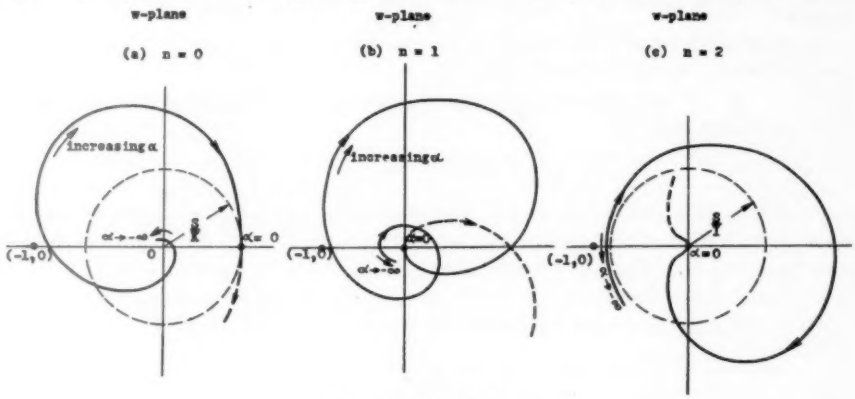

FIG. 1. $f(i\alpha)$ vs. α.

$-\infty \leqq \alpha \leqq 0$. As α increases from 0 to ∞, each graph traces the second branch of the spiral which is symmetrical to the first branch with respect to the real axis. Figure 2 shows typical graphs of $|f(i\alpha)|$ versus α for $n = 0, 1, 2$ and $n > 2$.

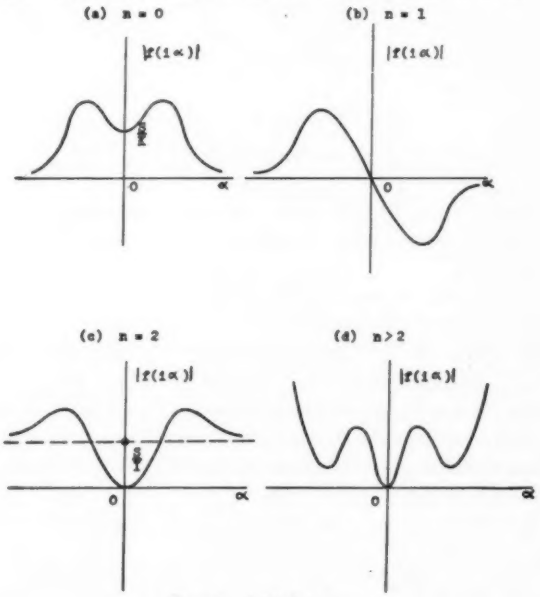

FIG. 2. $|f(i\alpha)|$ vs. α.

Now if $p \to \pm i\infty$, then

$$f(p) \to \frac{S}{I} p^{n-2} \qquad \text{and} \qquad |f(p)| \to \infty$$

except for $n = 0, 1, 2$. Thus, for $n > 2$, as p traces the imaginary axis, the graph of Eq. (9) will always enclose the point $(-1, 0)$. This means that (7) will always have roots in the right half-plane. The general stability criterion for Eq. (2) can now be formulated.

If the feedback term is proportional to the retarded derivative $y^{(n)}(t - \tau)$, the system will be unconditionally unstable whenever $n > 2$.

For $n = 0, 1, 2$ separate stability criteria have to be formulated in each case. The general procedure is to set

$$|f(i\alpha)| = 1 \tag{10}$$

and solve for α. In general, one obtains four real roots, two for each branch of the spiral, symmetric with respect to the real axis. It can be shown [5] that the angle β of the spiral is monotone in α and that each branch of the spiral will have a sector, say $\beta_1 \leqq \beta \leqq \beta_2$, within which $|f(i\alpha)| \geqq 1$. This is illustrated in Fig. 3. The stability criterion is now reduced to the condition that this sector does not enclose the negative real axis.

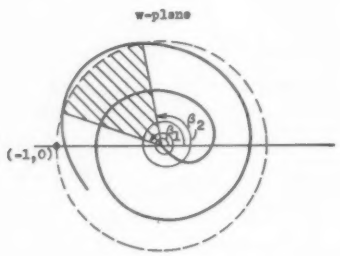

FIG. 3.

A Stable Configuration for $n = 1$.

In degenerate cases Eq. (10) may have only one real root for each branch of the spiral and the criterion is accordingly simplified. If (10) has no real roots, the system is unconditionally stable.

(Because of the symmetry of each spiral with respect to the real axis, it is sufficient to formulate criteria for one branch only).

A detailed application for the above procedure to each case of feedback will be found in reference [5]. Summarized below are the allowable ranges in τ for which the system will be stable. The expressions are written for the branch of the spiral on which $\alpha > 0$. We let α_1 and α_2 designate the two roots of (10) on this branch and ϕ_1 and ϕ_2 be the corresponding phase angles computed from (9c).

Feedback proportional to $y(t - \tau)$; $n = 0$.

(a) α_1, $\alpha_2 \neq 0$, real, $\alpha_1 > \alpha_2$. In this case $S < K$ and $2IK > R^2$. The range in τ is

$$\alpha_2^{-1}[\phi_2 + (2n - 1)\pi] < \tau < \alpha_1^{-1}[\phi_1 + (2n + 1)\pi], \tag{11}$$

where n is allowed integer positive values. Examination of (11) shows that for any given system (fixed α_1 , α_2 , ϕ_1 and ϕ_2) τ is allowed a band spectrum of values for increasing n with the band width given by (11).

(b) $\alpha_1 = \alpha_2 \neq 0$, real. Then $S \leqq K$ and the allowable values of τ are given by

$$\tau \neq \alpha^{-1}(n\pi + \phi), \tag{12}$$

where n is any odd integer. In this case the "instability" sector has degenerated into a straight line.

(c) $\alpha_1 > 0$, real, α_2 imaginary. Then $S > K$ and the range is

$$0 < \tau < \alpha_1^{-1}(\pi + \phi_1). \tag{13}$$

Here the sectors from the two branches have degenerated into one which is symmetrical with respect to the positive real axis.

(d) α_1 and α_2 both imaginary. This occurs if and only if

$$S < K \qquad \text{and} \qquad R^2 > 2IK.$$

The system is stable for all values of τ.

Feedback proportional to $y'(t - \tau)$; $n = 1$.

(a) α_1 , $\alpha_2 \neq 0$, real, $\alpha_1 > \alpha_2$. In this case $R < S$. Equation (11) applies.

(b) $\alpha_1 = \alpha_2 \neq 0$, real. A sufficient condition for this to happen is $R = S$. Equation (12) applies.

(c) α_1 , α_2 both imaginary. This occurs if and only if $R > S$. The system is stable for all values of τ.

Feedback proportional to $y''(t - \tau)$; $n = 2$.

As seen from Fig. 1(c) and Fig. 2(c) the system is *unconditionally unstable if* $S > I$.

(a) α_1 , $\alpha_2 \neq 0$, real, $\alpha_1 > \alpha_2$. Then $S < I$ and $R^2 < 2IK$. Equation (11) applies.

(b) $\alpha_1 = \alpha_2 \neq 0$, real. A sufficient condition is $S = I$ and $R^2 = 2IK$. Equation (12) applies.

(c) α_1 , α_2 both imaginary. This occurs if and only if

$$S < I \qquad \text{and} \qquad R^2 > 2IK$$

The system is stable for all values of τ.

Bibliography

[1] Seymore Sherman, *A note on stability calculations and time lag*, Q. Appl. Math. **5**, 92 (1947).

[2] N. Minorsky, *Self excited oscillations in dynamical systems*, J. Appl. Mech. **64**, A 65-71 (1942).

[3] M. F. Gardner and J. L. Barnes, *Transients in linear systems*, Wiley & Sons, New York, 1942.

[4] L. McColl, *Fundamental theory of servomechanisms*, Van Nostrand Co., New York, 1945.

[5] H. I. Ansoff, *Stability of linear oscillating systems with constant time lag*, this is a more detailed version of the present paper which has been submitted to the Journal of Applied Mechanics.

[6] A. Hurwitz, *Funktionentheorie*, J. Springer, Berlin, 1929, p. 106.

ERRORS IN SIMULTANEOUS LINEAR EQUATIONS[1]

By RAYMOND REDHEFFER (*Research Lab. of Electronics, M. I. T.*)

1. The problem. After computing the unknowns in a linear system $\sum_1^n a_{ij}\, x_i = b_j$ with non-vanishing determinant, one may check the work by substitution into the original equations. If the j-th equation is satisfied within an error e_j, it is natural to inquire into what can be said about the corresponding error in the unknowns. This question is the subject of the present note. Analytically we have

$$\sum a_{ij}(x_i + E_i) = b_j + e_j, \tag{1}$$

where x_i are the exact values of the unknowns, and where the errors E_i in the unknowns are to be estimated in terms of the substitution errors e_j. This problem was suggested by Prof. P. Franklin.

2. A solution. Regarding $\{E_i\}$ and $\{e_i\}$ as vectors, let us estimate the maximum length of one in terms of the length of the other; in other words let us find max $(\sum E_i^2)^{1/2}$ subject to $(\sum e_i^2)^{1/2} = e$. This is equivalent to finding the minimum of $\sum c_i^2$ subject to a constant value of $\sum E_i^2$, and such a problem lends itself to standard methods. Using the fact that the x_i satisfy the equations exactly, we have

$$\sum a_{ij}E_i = e_j\,. \tag{2}$$

We are thus required to minimize

$$\sum_j \left(\sum_i a_{ij}E_i \right)^2 \tag{3}$$

subject to the condition that $\sum E_i^2 =$ constant. By collecting terms in (3) we find that the matrix of the quadratic form is AA', where $A = (a_{ij})$ is the matrix of the original system and A' is its transpose. Hence, if e is the length of the known error vector, $e = (\sum e_i^2)^{1/2}$, and E is the length of the unknown error vector, $E = (\sum E_i^2)^{1/2}$, then we have[1]

$$E \leq e/(\lambda_m)^{1/2}, \tag{4}$$

where λ_m is the minimum characteristic value of the matrix AA'. When the matrix A is symmetrical, in particular, we have

$$E \leq e/|\,\lambda_m'\,|, \tag{5}$$

where λ_m' is (in absolute value) the minimum characteristic value of A. These results are the best possible, in that equality is always attained, for given e, with some set of values e_j.

3. Approximation. If λ_i are the characteristic values of AA' we know that

$$\sum \lambda_i = \sum\sum a_{ij}^2 = T, \tag{6}$$

the trace of AA'. Also

$$\Pi\lambda_i = |\,A\,|^2 \tag{7}$$

---

[1]Received Nov. 21, 1947. This work has been supported in part by the Signal Corps, the Air Materiel Command, and the Office of Naval Research.

[2]R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, vol. 1, J. Springer, Berlin, 1931, p. 21.

if $|A|$ stands for the determinant of A. By (3) it is clear that AA' is positive definite, so that $\lambda_i > 0$, and hence we may use (6) to find

$$\sum_{i \neq m} \lambda_i < T, \tag{8}$$

while in any case we have, by (7),

$$\lambda_m \prod_{i \neq m} \lambda_i = |A|^2. \tag{9}$$

If the sum of n positive quantities is constant the product is maximum when they are all equal. Hence we find, by (8) and (9),

$$\lambda_m > \frac{|A|^2}{T^{n-1}} (n-1)^{n-1} \tag{10}$$

with the final result

$$E < \frac{e}{||A||} \left(\frac{T}{n-1} \right)^{(n-1)/2} \tag{11}$$

which is not difficult to compute numerically.

From the derivation it is clear that this upper bound is not far from optimum, whenever λ_m is small compared to the sum of the λ_i, and in addition the other λ_i do not vary over too wide a range of values. In the case of any normal orthogonal transformation, for example, we have

$$E < \frac{e}{1} \left(\frac{n}{n-1} \right)^{(n-1)/2} \to e(2.718\cdots)^{1/2} \cong 1.65e$$

by (11), whereas the optimum inequality is $E \leq e$.

4. Numerical examples. Suppose that approximate values of the unknowns are substituted in a system with the matrix

$$A = \begin{pmatrix} 3 & -2 & -2 \\ -2 & 5 & 1 \\ -2 & 1 & 4 \end{pmatrix} \tag{12}$$

and are found to satisfy the equations with errors of 0.1, 0.1 and 0.2 respectively. To find the maximum possible error in any one unknown we proceed as follows. The determinant of the system A is 29, and for T we take the sum of the squares of the entries in (12), which gives $T = 68$. Substituting in (11) with $n = 3$, we find

$$E < 1.17e \tag{13}$$

In the present case $e = (.01 + .01 + .04)^{1/2} = .22$ so that E, and hence the maximum error in any one unknown, cannot exceed 0.26.

The characteristic values were found by J. G. Linvil to be 1.13, 7.39, and 3.49. Using $\lambda_m = 1.13$ in (5) we obtain the optimum inequality

$$E < 0.89e \tag{14}$$

which is only a slight improvement on (13).

BOOK REVIEWS

Engineering applications of fluid mechanics. By J. C. Hunsaker and B. G. Rightmire. McGraw-Hill Book Company, Inc., New York and London, 1947. ix + 494 pp. $5.00.

In preparing this book which obviously represents the fruit of a long teaching experience, the authors aimed at converting the "traditional course in hydraulics into a more fundamental treatment of the action of fluids generally. The authors were more anxious that the student understand flow phenomena than that he be familiar with details of many practical devices." Accordingly, the emphasis is on the basic ideas of modern fluid mechanics. The scope of this book is best indicated by the following chapter headings: Introductory survey—Statics—Kinematics and Continuity—Dynamics of an ideal fluid—Energy relations for steady flow—Dimensional analysis and similitude—Incompressible flow in closed circuits—Compressibility phenomena—Drag—Wing theory—Hydrodynamic lubrication—Boundary lubrication—Hydraulic turbines—Pumps, fans and compressors—Propellers and jets—Fluid couplings and torque converters—Hydraulic transmissions and controls. The book is clearly written and constitutes an excellent introduction into fluid mechanics.

W. PRAGER

The physical principles of wave guide transmission and antenna systems. By W. H. Watson. Clarendon Press, Oxford, 1947. xiii + 207 pp., 95 text-figures. $7.00.

This new book deals extensively with the systems' aspects of microwave transmission, which in other books are treated but briefly if at all. Wave propagation in uniform wave guides is discussed briefly but sufficiently for the main purpose of the book. Expressions for the impedances and admittances of slots in every conceivable position and for some other wave guide elements are given without derivation and the reader is referred to the original sources for the details. Omission of these details is in keeping with the main object of the book, which is to discuss systems of wave guide elements rather than the elements themselves. It also enables the author to give more space to discussion of the results and to concentrate an amazing amount of information into a small book. Wave guide elements, however, are not altogether neglected; the last chapter is devoted to the methods of their analysis.

At those frequencies for which the energy in a wave guide is transmitted to large distances only by the dominant wave, the wave guide discontinuities (probes, slots, etc.) may be represented either as series impedances or shunt susceptances or, more generally, as T or Π passive networks. They give rise to reflected waves which, when superimposed on the original wave system, produce a new wave system. The transformation from one wave system to the other may conveniently be expressed by a matrix. The transformation of a given wave system by any number of discontinuities may thus be expressed by suitable operations on matrices. These basic ideas are presented in Chapter I and are illustrated by their application first to a strip transmission line consisting of two parallel conducting strips and then to waves in free space.

Chapter II is a summary of the essential aspects of dominant waves in rectangular wave guides; and Chapter III is devoted to measurements and to the application to wave guides of the ideas introduced in Chapter I. Higher modes of propagation are discussed briefly in Chapter IV. Chapter V includes a discussion of "Babinet's" principle as applied to wave guides and its applications. Important wave guide discontinuities such as irises, probes, bends and twists are also treated in this chapter. Coupling of a wave guide to free space and to another wave guide by means of slots is studied in the following two chapters. Wave guide arrays are taken up in Chapter VIII. In Chapter IX the author discusses briefly a number of miscellaneous microwave devices. Then, in Chapter X the author suddenly recalls (the responsibility for this opinion is strictly ours) that he is a mathematician and that in accordance with popular beliefs a mathematician's book should contain more formidable mathematical equations; hence he proceeds to introduce some of the mathematical methods which are required in the theoretical studies of wave guide discontinuities.

It is safe to assume that this book will be welcomed by all who are interested either in the theory or in the practice of microwave transmission. It does not attempt to compete with other books in this field, and it has no competitors.

S. A. SCHELKUNOFF

MECHANICS. New 2nd edition

, Carnegie Institute of Technology, and H. A. Gibbons, University of Toronto. Ready

comprehensively from the fundamentals of assumptions in geometric theory, Lagrange's equa- of constraint, Newtonian electron equations, Lorentz method, Gaussian's principle of concepts, canonical theory, space, time, reference systems, etc. Many changes have the clarity of the treatment.

METHODS IN ENGINEERING

n Jacobs, California Institute of Technology, and Maurice A. Biot, Columbia Uni- 1950

mathematical treatment of engineering problems. Necessary tools beyond elementary ; it drives concretion with practical problems from the fields of mechanical, civil, and

THEORY OF ELASTICITY

, University of California at Los Angeles. 373 pages, $4.50

mathematical treatment of the mechanics of deformable media. After giving a thorough classical theory of elasticity, the author deals with the application of the theory to , and flexure of isotropic cylindrical bodies and with variational methods in the

Send for copies on approval

McGRAW-HILL BOOK COMPANY, Inc.

New York 18, N. Y.

. 215

. 221

. 257

. 267

. 279

. 301

. 320

. 331

. 351

. 357

. 373

. 374

. . . . . a most-widely reviewed text in theory of equations

OF THE THEORY OF EQUATIONS. by the late J. V. Uspensky, Stanford . . . 352 pages, $4.50. An up-to-date, thorough, explicit treatment of the subject, development, emphasizing both theory and numerical methods. The sections on . . . round methods are particularly well worthy.

Send for a copy on approval

McGraw-Hill Book Company, Inc.

. . . 42nd Street, New York 18, N. Y.